

Universidad Autónoma de Madrid



Facultad de Ciencias

Departamento de Biología Molecular

**La red global de biodegradación: aproximación desde la
biología de sistemas y la biología sintética.**

Memoria presentada para optar al grado de Doctor en Ciencias por:

Almudena Trigo Lorenzo

Tutor / Director:
Alfonso Valencia

Director:
Ildefonso Cases

SUMMARY

Human activities produce a remarkable amount of compounds, many of them harmful to the natural environment. Microbial communities have developed a natural capacity to degrade recalcitrant compounds and so to clean-up polluted areas. Based on this potential, humans have tried to repair the environment. However the results have been not as good as expected, implying the necessity for a better knowledge of the biodegradation processes.

In an attempt to understand this process better, first I have associated protein sequences to a large number of reactions in the network allowing the study of the topology, organization and the evolution of the global biodegradation network from a systems biology point of view, and second I, designed and modeled a synthetic bioremediation system for arsenic.

The results show that the global biodegradation network is scale-free and small-world like the Central Metabolism Network, but unlike this, the biodegradation network is not hierarchical and has a funnel structure. Also this analysis reveals an increase in complexity and the promiscuity of the enzymatic complexes more distant to the Central Metabolism, and an increase in the number of homologs with the same specificity of the proteins in this zone. In the same manner, this work shows that it is not possible to detect a close evolutionary relation by sequence comparison between the majority of the biodegradation proteins and those of the Central Metabolism Network. The quantification of the evolutionary mechanisms in the biodegradation network shows that the driving force behind it is the same as the Central Metabolism, the patchwork model, followed by the pathway duplication model and rarely the retro-evolution. Also, each one of these theories has a tendency in the type of duplication, patchwork and retro-evolution gene duplication and pathway duplication functional duality. Finally, using my background as engineer and the learned knowledge about biodegradation, I designed modeled and simulated a regulatory circuit controlled by arsenic concentration which has the potential for bioremediation of arsenic pollution in soil

ÍNDICE

1	INTRODUCCIÓN	19
1.1	Biodegradación.	21
1.2	Biorremediación.	22
1.3	Nuevas disciplinas.....	24
1.3.1	Biología de sistemas.	25
1.3.1.1	Topología de las redes.	26
1.3.1.1.1	Teoría: Estudio de la topología de las redes.	26
1.3.1.1.1.1	Redes “libre de escala”.	28
1.3.1.1.1.2	Redes de “mundo pequeño”.	28
1.3.1.1.1.3	Jerarquía y modularidad en las redes.....	30
1.3.1.1.2	Topología de la red del Metabolismo Central.	31
1.3.1.1.3	Antecedentes en el estudio de la topología de la red de Biodegradación.	32
1.3.1.2	Evolución de las redes.	34
1.3.1.2.1	Teoría: Formación y evolución de las redes.	34
1.3.1.2.2	Evolución: Red del Metabolismo Central.	36
1.3.1.2.3	Antecedentes en el estudio de la evolución en la red de biodegradación.....	37
1.3.2	Biología Sintética.	38
1.3.2.1	Contexto.	38
1.3.2.2	Aplicación de la biología sintética en biodegradación.	41
2	OBJETIVOS	43
2.1	Objetivo General.	45
2.2	Objetivos Específicos.....	45
3	MATERIALES Y MÉTODOS.....	47
3.1	Base de datos BioNeMo.....	49
3.1.1	Asociación de cada nodo de la red (reacción) con su enzima.	49
3.1.2	Realización de la base de datos.	49
3.1.3	Realización del servidor web.....	566

3.2	Análisis del nivel de descripción de la red de Biodegradación.	57
3.2.1	Formación de la red de biodegradación.....	57
3.2.2	Comparación de la red de Biodegradación con la red del Metabolismo Central.	58
3.2.3	Comparación de las reacciones que llegan al Metabolismo Central de las que no.	59
3.2.4	Distancia al Metabolismo Central.	59
3.3	Topología de la red de Biodegradación.	60
3.3.1	Red libre de escala.	60
3.3.2	Red “mundo pequeño”	60
3.3.3	Red jerárquica.....	61
3.4	Organización de la red de Biodegradación.	62
3.4.1	Complejidad.	62
3.4.2	Especificidad.	63
3.4.3	Variabilidad.	63
3.5	Evolución de la red de Biodegradación.	64
3.5.1	Dominios funcionales de la red de Biodegradación y la del Metabolismo Central.	64
3.5.2	Evolución de las proteínas.....	64
3.5.3	Evolución de las reacciones.....	65
3.5.4	Bloques de reacciones consecutivas.	66
3.5.5	Teorías de evolución de las redes metabólicas.....	67
3.5.6	Tipos de duplicación entre pares de reacciones relacionadas por similitud de secuencia.	68
3.6	Biología Sintética: El detoxificador de arsénico.	69
4	RESULTADOS	73
4.1	Conceptos importantes de la red de Reacciones de Biodegradación.	75
4.2	La base de datos BioNeMo y sus herramientas.	78
4.2.1	Base de Datos.	78
4.2.2	Interfaz de programación de aplicaciones (API).	81
4.2.3	Servidor Web.....	82
4.3	Grado de descripción de la red de Biodegradación.....	90

4.3.1	La red de Biodegradación Vs la red del Metabolismo Central.	92
4.3.2	Biodegradación total Vs. biodegradación parcial	93
4.3.3	Caracterización a lo largo de la red de Biodegradación.	94
4.4	Estudio topológico de la red de Biodegradación.....	97
4.4.1	Estructura libre de escala.	97
4.4.2	Red de “mundo pequeño”.....	100
4.4.3	Jerarquía de la red.....	101
4.5	Organización y propiedades de las reacciones en la red de Biodegradación.	106
4.5.1	Complejidad de las metaenzimas encargadas de las reacciones.	106
4.5.2	Especificidad de las metaenzimas que realizan las reacciones.	110
4.5.3	Variabilidad de las proteínas.	112
4.6	Formación y evolución de la red de Biodegradación.	116
4.6.1	Comparación funcional entre Biodegradación y el Metabolismo Central.	117
4.6.2	Evolución de las proteínas de biodegradación.	119
4.6.3	Evolución de las reacciones de biodegradación.	121
4.6.4	Evolución del ensamblaje de las reacciones de la red.	124
4.6.5	Teorías de Evolución.	126
4.7	Aplicación de la biología sintética: el detoxificador de arsénico.	132
4.7.1	Diseño del detoxificación.	134
4.7.2	Modelado del detoxificador.....	139
4.7.3	Simulación del detoxificador.....	141
4.7.3.1	Primer módulo: Sensor de arsénico y metilación.	142
4.7.3.2	Segundo módulo: Represión de CI.....	146
4.7.3.3	Tercer módulo: Activación de la bomba de arsénico.	147
4.7.3.4	Funcionamiento global del sistema.	149
4.7.3.5	Posibles configuraciones que mejorarían el sistema. .	152
5	DISCUSIÓN.....	155
5.1	Aplicación de la Biología de sistemas en biodegradación: topología, organización y evolución.	157
5.1.1	Base de datos BioNeMo.	157

5.1.2	Análisis del nivel de descripción de la red de Biodegradación.	158
5.1.3	Topología de la red de Biodegradación.....	158
5.1.4	Organización de la red de Biodegradación.....	160
5.1.5	Evolución de la red de Biodegradación.....	161
5.2	Teoría sobre la formación y evolución de la red de biodegradación.....	165
6	CONCLUSIONES	167
7	BIBLIOGRAFÍA	171
8	APÉNDICE 1	185
9	APENDICE 2	189
10	APÉNDICE 3	195
11	APÉNDICE 4	195

Índice de Figuras.

Figura 1. Modelos de redes y sus propiedades topológicas.....	27
Figura 2. Recableado.....	29
Figura 3. Teorías evolutivas en redes metabólicas.....	34
Figura 4. Sistemas de autoregulación.....	40
Figura 5. Estructura de la información "enzimática" en la base de datos BioNeMo.....	50
Figura 6. Transformación: red de compuestos a red de reacciones.....	57
Figura 7. Conceptos enzimáticos y sus relaciones.....	76
Figura 8. Relaciones en BioNeMo.....	80
Figura 9. Estructura del servidor web de BioNeMo.....	83
Figura 10. Página de inicio del servidor web BioNeMo.....	84
Figura 11. Página de resultados "google" (Búsqueda por <i>Benzoate</i>)..	85
Figura 12. Página de ruta de Benzoato.....	86
Figura 13. Página de reacciones.....	87
Figura 14. Página gen/proteína.....	88
Figura 15. Definición de los niveles de caracterización de las reacciones.....	90
Figura 16. Comparación del grado de descripción de las reacciones en la red de Biodegradación y en la del Metabolismo Central clasificada en niveles.....	93
Figura 17. Comparación del grado de descripción de las reacciones que forman parte de procesos de biodegradación total respecto a las reacciones que participan en biodegradación parcial.....	94

Figura 18. Grado de descripción de las reacciones de la red de biodegradación que participan en procesos de biodegradación total, clasificados según el nivel de caracterización y la distancia de las reacciones a la entrada a la red del Metabolismo Central	95
Figura 19. Visión global de la red de reacciones de biodegradación..	96
Figura 20. Distribución libre de escala de la conectividad de entrada.	98
Figura 21. Distribución libre de escala de la conectividad de salida...	99
Figura 22. Distribución libre de escala de la conectividad total.....	99
Figura 23. Coeficiente de Clustering en función del número de conexiones de las reacciones..	102
Figura 24. Relación entre el número de conexiones de entrada y salida de las reacciones de la red de biodegradación.....	103
Figura 25. Distribución del número de conexiones de entrada respecto a la distancia de la reacción a la entrada al Metabolismo Central.....	104
Figura 26. Distribución del número de conexiones de salida respecto a la distancia de la reacción a la entrada al Metabolismo Central.....	104
Figura 27. Distribución del número de conexiones totales respecto a la distancia de la reacción a la entrada al Metabolismo Central.....	105
Figura 28. Distribución de la complejidad de las reacciones respecto a la distancia al Metabolismo Central.	107
Figura 29. Distribución respecto a la distancia a la entrada al Metabolismo Central del número de dominios por proteína.	108
Figura 30. Distribución del número de subunidades por complejo de proteínas respecto a la distancia a la entrada al Metabolismo Central.	109
Figura 31. Ejemplo complejidad en la red de Biodegradación.	110

Figura 32. Distribución de la especificidad de los complejos de las reacciones en biodegradación respecto a la distancia a la entrada al Metabolismo Central.	111
Figura 33. Ejemplo especificidad en la red de biodegradación.....	112
Figura 34. Distribución del número de homólogos de las proteínas de la red de biodegradación, respecto a su distancia a la entrada del Metabolismo Central, utilizando un porcentaje de identidad entre secuencias homólogas superior al 95%.	113
Figura 35. Distribución del número de homólogos de las proteínas de la red de biodegradación, respecto a su distancia a la entrada del Metabolismo Central, utilizando un porcentaje de identidad entre secuencias homólogas superior al 50%.	114
Figura 36. Frecuencia de los dominios funcionales de biodegradación respecto a su frecuencia en la red del Metabolismo Central.	118
Figura 37. Relación evolutiva detectable mediante homología de secuencias de las proteínas de biodegradación con las del Metabolismo Central.	120
Figura 38. Relación evolutiva detectable mediante homología de secuencia de los complejos de biodegradación en la red del Metabolismo Central.	121
Figura 39. Distribución de los tipos de reacciones (CM, no CM y mixtas) respecto a su distancia a la entrada del Metabolismo Central.	123
Figura 40. Ejemplos del origen de las reacciones mixtas.....	123
Figura 41. Pares de reacciones consecutivas en biodegradación cuyos homólogos también lo son en la red del Metabolismo Central.	124
Figura 42. Ejemplo de bloques de reacciones consecutivas en biodegradación y el Metabolismo Central.....	125
Figura 43. Conceptos y tipos de eventos evolutivos.	127

Figura 44. Representación de las teorías evolutivas en la red de biodegradación.	128
Figura 45. Distribución de las teorías evolutivas a lo largo de la red de biodegradación.	129
Figura 46. Representación de cada tipo de RRSS en las teorías evolutivas.	130
Figura 47. Ejemplos tipos de RRSS en la red de biodegradación.	131
Figura 48. Descripción del operón Ars.	133
Figura 49. La representación electrónica de cada uno de las partes del circuito.	135
Figura 50. Esta figura muestra el diseño electrónico que se comportaría de la manera mostrada en la Figura 49.	136
Figura 51. Esquema del funcionamiento del detoxificador.	138
Figura 52. Diagrama del circuito biológico del detoxificador de arsénico.	139
Figura 53. Representación gráfica de las reacciones del circuito global.	140
Figura 54. Modularización del sistema.	141
Figura 55. Representación de las reacciones del primer módulo.	143
Figura 56. Función de transferencia del primer módulo con los parámetros por defecto.	144
Figura 57. Función de transferencia del primer módulo tras aumentar la tasa de traducción de Cyt19.	145
Figura 58. Función de transferencia del primer módulo tras disminuir la tasa de degradación de la proteína Cyt19.	145
Figura 59. Representación de las reacciones del segundo módulo. ..	146

Figura 60. Función de transferencia del segundo módulo con los parámetros por defecto.	147
Figura 61. Representación de las reacciones del tercer módulo.....	148
Figura 62. Función de transferencia del tercer módulo con los parámetros por defecto.	149
Figura 63. Función de transferencia del sistema global tras disminuir la tasa de degradación de la proteína Cyt19.	150
Figura 64. Función de transferencia del sistema global tras disminuir la tasa de degradación de la proteína Cyt19, disminuir la tasa de degradación de la proteína ArsB y aumentar la tasa de traducción de ArsB.....	151
Figura 65. Análisis de los retrasos en la respuesta del sistema global. Variación de las entradas y salidas de los módulos del sistema frente a la variación de arsénico.	152
Figura 66. Configuración: mejora en la velocidad de respuesta del sistema. La construcción se diferencia del circuito original en utilizar un activador (araC) en lugar de dos represores en serie (LacI y CI). .	153
Figura 67. Configuración: mejora en el nivel de activación de cada mecanismo.....	153
Figura 68. Configuración: mejora en el nivel de activación de cada mecanismo y en la respuesta del sistema de bombeo.....	154

Glosario

ADN: Ácido Desoxirribonucleico.

API: del término inglés *Application Program Interface*, interfaz de programación de aplicaciones.

BioNeMo: *Biodegradation Network Modelling*.

BTEX: hidrocarburos aromáticos: **B**enceno, **T**olueno, **E**tilbenceno y **X**ileno.

CGI: del inglés *Common Gateway Interface.*, interfaz de entrada común.

CM: del inglés *Central Metabolism*, se traduce como Metabolismo Central.

C02: Dióxido de Carbono.

Coefficiente de *clustering*: indica la probabilidad de que dos nodos conectados directamente a un tercer nodo, estén conectados entre sí.

Conectividad de un nodo: número de enlaces que le conectan con otros nodos de la red.

Diámetro de una red: o distancia media entre nodos se define como el tamaño medio de los caminos mínimos entre todos los pares de nodos.

DDT : 2,2-bis-(p-clorofenil)-1,1,1-tricloroetano

Distancia entre nodos: número mínimo de enlaces que hay que atravesar para alcanzar desde el nodo origen el nodo destino.

DNT: 2,4-Dinitrotolueno.

Dualidad funcional: Cuando las reacciones se relacionan por similitud de secuencia (RRSS) debido a que son realizadas por la misma metaenzima.

Duplicación genética: Cuando las reacciones se relacionan por similitud de secuencia (RRSS) debido a que son realizadas por metaenzimas homólogas pero no la misma.

EGFP: del inglés *Enhanced Green Fluorescent Protein*, proteína verde fluorescente.

Estructuras 3D: Estructuras tridimensionales

FAD: Flavín Adenín Dinucleótido

GenBank: Base de datos de secuencias del *National Institutes of Health* (NIH) de Estados Unidos.

Grafo: Conjunto de nodos y enlaces que permite representar relaciones entre elementos.

GFP: del inglés *Green Fluorescent Protein*, proteína verde fluorescente.

HGT: del inglés *Horizontal Gene Transfer*, transferencia horizontal de genes.

HTML: del inglés *HyperText Markup Language*, lenguaje de etiquetas de hipertexto.

Hub: nodos con alta conectividad (muchas conexiones).

Isoenzima: enzimas que difieren en la secuencia de aminoácidos, pero que catalizan la misma reacción química.

KEGG: del inglés *Kyoto Encyclopedia of Genes and Genomes*, es una base de datos con una colección manual de rutas metabólicas.

Matlab: del inglés *Matrix Laboratory*, es un programa de análisis numérico creado por “The MathWorks”.

Metaenzima: concepto abstracto que englobaría todos los complejos de proteínas distintos capaces de realizar la misma reacción.

MGEs: del inglés *Mobile Genetic Elements*, elementos genéticos móviles.

NAD: Nicotinamida Adenín Dinucleótido

NCBI: *National Center for Biotechnological Information.*

Número EC: *Enzyme Commission number*, esquema de clasificación numérica para las enzimas basado en las reacciones químicas que realizan.

OGMs: Organismos Genéticamente Modificados.

Operón: grupo de genes cuya expresión está regulada por los mismos elementos de control (promotor y operador) y genes reguladores.

PAHs: del inglés *Polycyclic Aromatic Hydrocarbons*, Hidrocarburos aromáticos policíclicos.

PCBs: del inglés *Polychlorinated Biphenyls*, policlorobifenilos

PCE: Percloroetileno

PDB: Del inglés *Protein Data Bank*. Es una base de datos sobre estructura de las proteínas y ácidos nucleicos.

Perl: Lenguaje Práctico para la Extracción e Informe.

Pfam: Base de datos compuesta por los perfiles obtenidos para distintos dominios o regiones conservadas de proteínas. Representan algunas estructuras conservadas evolutivamente que tienen una implicación con la función de la proteína.

pH: potencial Hidrógeno.

PostgreSQL: es un servidor de base de datos objeto relacional libre bajo la licencia de *Berkeley Software Distribution*.

Reacción: transformación que convierte un sustrato en un producto, y al cuál se le puede asignar una metaenzima. complejo de proteínas encargadas de llevar a cabo tal transformación.

RRSS: Reacciones relacionadas por similitud de secuencia.

SQL: Lenguaje de Consulta Estructurado. Lenguaje de acceso a bases de datos relacionales.

Swissprot: es una base de datos biológica de secuencias de proteínas anotadas manualmente.

TCE: Tricloroetileno

TNT: Trinitrotolueno.

Trembl: Del inglés *Translation of EMBL* es una base de datos biológica de secuencias de proteínas anotadas de forma automática.

UMBBD: del inglés *University of Minnesota Biocatalysis/Biodegradation Database*, es una base de datos de rutas de degradación de compuestos recalcitrantes

Uniprot: Base de datos que engloba la información de Swissprot y la de Trembl.

XFP: Proteína fluorescente.

1 INTRODUCCIÓN

1.1 Biodegradación.

La biodegradación es la capacidad metabólica de los microorganismos para transformar compuestos orgánicos en otros más simples que pueden integrarse en los ciclos biogeoquímicos naturales. Desde hace alrededor de 100 años la actividad industrial se ha incrementado de manera considerable. Este hecho ha provocado el aumento de la concentración final de compuestos tóxicos que de manera natural, pero en menores cantidades, se encuentran en la naturaleza y de la aparición de nuevos compuestos contaminantes sintetizados artificialmente (Díaz, 2004).

En algunas ocasiones la liberación de dichos compuestos en el medio ambiente ha producido efectos nocivos inicialmente desconocidos. Este es el caso del insecticida de 2º generación DDT (2,2-bis-(p-clorofenil)-1,1,1-tricloroetano) (Beard, 2006), el cual pasó de ser un benefactor de la humanidad a enemigo público entre los años 1970 a 80 y con ello llegó su prohibición. De este modo, aunque en 1948, el químico suizo Müller recibió el Premio Nobel en reconocimiento al impresionante avance que este producto había representado en la lucha contra las enfermedades y las plagas, más tarde se fueron descubriendo importantes problemas asociados a su uso, como su lenta conversión a sustancias no tóxicas y su poca solubilidad en agua, lo que evita que se elimine en la orina, permaneciendo en los tejidos de los organismos y acumulándose a lo largo de la cadena trófica.

En otras ocasiones, el daño se ha producido por una gran acumulación de residuos industriales o por vertidos accidentales. En 2002, el petrolero “Prestige” vertió miles de toneladas de fuel frente a las costas gallegas produciendo una gran marea negra. Las consecuencias de este desastre ecológico no sólo afectaron a la vida marina y al ecosistema terrestre, sino que las pérdidas económicas asociadas a la actividad de esas zonas costeras fueron considerables (Suris-Regueiro et al, 2007).

La estabilidad de la estructura química de estos compuestos contaminantes es la razón principal de su persistencia en la biosfera sin alterarse y por lo que se les denomina compuestos recalcitrantes. Ante la presencia de estos compuestos en el medio, los microorganismos han adaptado su comportamiento extendiendo sus capacidades metabólicas

para sobrevivir, convirtiendo estos compuestos en otros menos tóxicos, u obtener energía, gracias a la presencia de carbonos e hidrógenos en las estructuras de estos compuestos.

Los microorganismos dotados genéticamente para la degradación pueden utilizar su potencial enzimático para transformar los contaminantes (biodegradar completamente hasta CO_2) o bien, simplemente degradarlos hasta productos intermedios, ya sea en un ambiente aerobio o anaerobio. El recorrido, catalizado por las enzimas microbianas a través de los diversos subproductos, se conoce como ruta de degradación. Esta atenuación natural (Rosenberg, 1996), aquella que se lleva a cabo por los microorganismos autóctonos, puede verse limitada por la carencia de nutrientes esenciales, la ausencia de aceptores adecuados de electrones y la inexistencia de condiciones medioambientales apropiadas (pH, potencial redox, humedad, temperatura). También la ausencia de poblaciones microbianas con el potencial enzimático adecuado para degradar los contaminantes, es decir, sin la adecuada inducción del operón/es o sin las enzimas catabólicas necesarias que reconozcan a los compuestos contaminantes de interés como sustratos también limita la atenuación natural. Por último, la presencia de componentes tóxicos en la mezcla contaminante que causen estrés en la célula pueden evitar que se produzca dicha atenuación (King, 1997).

1.2 Biorremediación.

Basándose en esta adaptación y potencial metabólico, el hombre ha tratado de subsanar los daños ocasionados en la naturaleza utilizando los microorganismos. A este proceso se le denomina biorremediación y trata de transformar contaminantes orgánicos en compuestos más simples poco o nada contaminantes.

La biorremediación puede aplicarse para eliminar cualquier compuesto que los microorganismos puedan captar o absorber, como son los hidrocarburos de todo tipo (alifáticos, aromáticos, BTEX, PAHs, ...), hidrocarburos clorados (PCBs, TCE, PCE, pesticidas, herbicidas, ...), compuestos nitroaromáticos (TNT, ...), y otros compuestos como organofosforados, cianuros, etc (Alexander, 2001). Es de gran utilidad en situaciones donde la degradación “in situ” del contaminante sea necesaria. Estas situaciones pueden producirse cuando la zona contaminada es de difícil acceso, y por tanto otros métodos como la

excavación no pueden utilizarse, o cuando no es posible retirar el vertido desde la zona contaminada, como es el caso de los numerosos escapes de petróleo que se producen en el mar.

La aplicabilidad de esta técnica depende de varios factores (Cases & de Lorenzo, 2005). Por una parte depende de las propiedades del contaminante, su biodegradabilidad; mientras en general los hidrocarburos alifáticos se degradan rápidamente, las cadenas ramificadas, dobles enlaces, sustituciones químicas y anillos bencénicos hacen a la molécula más resistente. Por otro lado también depende de la presencia de comunidades microbianas con capacidad enzimática para metabolizar el compuesto. La biorremediación también es función de la disponibilidad del contaminantes; para que la degradación de un contaminante se pueda producir, es necesario que interaccione con la célula. Por último depende de las condiciones del medio contaminado (nutrientes, pH, temperatura, etc), las cuales permiten o limitan el crecimiento microbiano y el metabolismo del compuesto.

Para salvar las limitaciones mencionadas en la atenuación natural de los microorganismos, en biorremediación se han utilizado las siguientes aproximaciones. Por una parte la bioaumentación consiste en la introducción de microorganismos específicos para mejorar la biodegradación (Walter, 1997; Atlas, 1999). Los microorganismos añadidos deben sobrevivir a los depredadores y competir con éxito con la población autóctona antes de ocupar los nichos potenciales. Por otro lado, la bioestimulación consiste la introducción de modificaciones en el medio, mediante el aporte de nutrientes, aireación y otros procesos (ej. cambios en el pH). Esta aproximación será válida siempre cuando los microorganismos autóctonos sean capaces de degradar el contaminante tras un proceso más o menos corto de aclimatación.

Haciendo uso de la bioaumentación, se comenzó a investigar como modificar genéticamente a los organismos para poder dotarlos de las capacidades degradativas necesarias (Timmis & Pieper, 1999; de Lorenzo, 2001).

A finales del 1970 y comienzo de los años 80, los genes de bacterias que codifican enzimas que catabolizan compuestos recalcitrantes empezaron a ser clonados y caracterizados. Estos esfuerzos guiaron a que en 1981 Gunsalus y Chakrabarty patentaran una cepa de *Pseudomonas* capaz de degradar camfor, octano, salicilato y naftaleno

[US Patent #425944]. Del mismo modo, a finales de los 80 Timmis demostró la posibilidad de incorporar capacidades en microorganismo, usando técnicas de DNA recombinantes con material genético de diferentes organismos y situando el conjunto de genes resultante bajo el control de promotores suficientemente fuertes (Ramos *et al*, 1987; Rojo *et al*, 1987). Siguiendo esta metodología consiguieron una cepa de *Pseudomonas* que eliminaba compuestos tan recalcitrantes como son los clorobenzoatos o alquil-benzoatos.

La utilización de estos organismos genéticamente modificados (OGMs) levantó un gran debate acerca de los posibles riesgos ecológicos (Lindow *et al*, 1989). Para evitarlos y controlar su utilización se exigió minimizar la transferencia lateral de los genes clonados, poder hacer un seguimiento de los genes y cepas modificadas y programar la muerte celular una vez que la biocatálisis ha cumplido su función o que los genes modificados se transfieran accidentalmente (Díaz *et al*, 1994; Ramos *et al*, 1994; Timmis & Pieper, 1999).

Aunque la ingeniería genética ha producido numerosas cepas capaces de degradar contaminantes intratables de otra forma, en una placa Petri o en un bioreactor (Reineke, 1998; Mishra *et al*, 2001), su aplicación “in situ” ha producido resultados bastante pobres (Sayler & Ripp, 2000; Díaz, 2004). Lo que se ha visto es que los OGMs son menos eficientes que sus homólogos naturales no manipulados (von Canstein *et al*, 1999; Strong *et al*, 2000). Esta ineficiencia es debida a la falta de conocimiento entorno al proceso de biodegradación, por lo tanto antes de poder utilizar la biorremediación de manera realista es necesario conocer mejor las reglas generales detrás de los procesos de biodegradación.

1.3 Nuevas disciplinas.

Los resultados anteriores muestran la necesidad de aplicar nuevas metodologías tanto para obtener más información sobre los procesos de biodegradación como para producir nuevas aplicaciones en biorremediación.

Aunque en un principio la modificación genética pareció ser la respuesta para solucionar contaminaciones ambientales, se ha visto que los procesos de biodegradación están enmarcados en una compleja red

de interacciones metabólicas y de regulación, difíciles de estudiar con las aproximaciones reduccionistas (Cases & de Lorenzo, 2005). La reciente aparición de las tecnologías “omics” (genómica, proteómica, metabolómica) y la aplicación en biología de la teoría de redes han dado lugar a la denominada biología de sistemas. Ésta disciplina abre una nueva perspectiva en el análisis de la red de biodegradación que podría proporcionar nuevos caminos por donde avanzar en biorremediación.

Por otra parte, la “recién nacida” biología sintética, aprovechando métodos y procedimientos de disciplinas tales como la biología molecular, la ingeniería genética, la ingeniería metabólica y la biología de sistemas, trata de introducir nuevos circuitos biológicos en los microorganismos como remodelar los existentes, empleando un proceso de estandarización análogo al utilizado en la ingeniería electrónica. Este campo presenta mucho potencial a la hora de diseñar e implementar nuevas aplicaciones en biorremediación.

1.3.1 Biología de sistemas.

La biología de sistemas es un área de investigación científica que estudia las complejas interacciones de los procesos biológicos usando un enfoque global (Kitano, 2001). Este campo comenzó a desarrollarse en los años sesenta, pero hasta el año 2000 no se ha producido su institucionalización académica. La idea en la que se basa la biología de sistemas no es nueva. Los científicos ya sabían que el estudio detallado de proteínas individuales (reduccionismo) era solo el primer paso hacia la comprensión del proceso completo de la vida. Sin embargo, los procedimientos experimentales disponibles solo permitían analizar una proteína cada vez (Cases & de Lorenzo, 2005).

En biología de sistemas no se divide el sistema en partes para estudiarlas una a una y así comprender su comportamiento global, sino que se estudia las interacciones entre las partes y cómo esas interacciones dan lugar a la función y comportamiento del sistema (Noble, 2006; Sauer *et al*, 2007). La visión de los procesos biológicos como sistemas ha permitido por una parte aplicar la teoría de redes para estudiar sus propiedades funcionales y por otra parte estudiar los mecanismos que participan en la formación y expansión de estos sistemas biológicos.

1.3.1.1 Topología de las redes.

1.3.1.1.1 Teoría: Estudio de la topología de las redes.

Durante los últimos años se ha desarrollado una activa investigación para explicar y comprender la estructura de las conexiones entre los elementos de diversos sistemas sociales (e.g. redes de colaboración científica o la *World Wide Web*), biológicos (e.g. redes de regulación entre genes o metabólicas), tecnológicos (e.g. la red de conexiones entre *routers*), etc ... En estos sistemas cada elemento se representa por un nodo y cuando entre dos nodos existe una relación (que dependerá del sistema) ésta se representa por una conexión entre ambos.

La Topología de una red es la descripción de la forma en la que se conectan sus nodos. Como explica Albertz-Làzló Barabasi en *Linked* (Barabasi & Bonabeau, 2003), los grafos y las redes tienen propiedades, debidas a su estructura, que explican el funcionamiento del sistema que componen. Por lo tanto, el conocimiento de la topología de la red haciendo uso de la teoría de redes es un paso clave a la hora de interpretar el funcionamiento, organización y evolución de la misma.

En biología de sistemas, muchos de los trabajos realizados se han centrado en el análisis topológico de los sistemas. El estudio de la topología está basado en el análisis de una serie de parámetros que dan una idea acerca de la estructura del sistema. Entre ellos se encuentra la conectividad de un nodo, definida como el número de enlaces que le conectan con otros nodos de la red. Si la red es dirigida, es decir, los enlaces tienen una dirección (que el nodo A esté conectado al nodo B, no implica que el nodo B esté conectado con el nodo A) se puede diferenciar entre conectividad de entrada (si sólo se tienen en cuenta los enlaces que entran en el nodo), conectividad de salida (si solo se estudian los nodos que salen del nodo) y conectividad total (si se tienen en cuenta todos los enlaces del nodo sin distinciones). Otro parámetro es el coeficiente de clustering entendido como la probabilidad de que dos nodos conectados directamente a un tercer nodo, estén conectados entre sí. La distancia entre dos nodos A y B, se medirá como el número de enlaces que hay que atravesar para alcanzar desde el nodo origen A el nodo destino B. De esta forma, se define la distancia media entre nodos o diámetro de una red como el tamaño medio de los caminos

mínimos entre todos los pares de nodos. Estos parámetros han sido ampliamente estudiados tanto en las redes regulares, en las cuales todos los nodos tienen el mismo número de vecinos (el mismo grado de conectividad), como en las redes aleatorias donde los nodos están conectados al azar (Figura 1Aa, 1Ab).

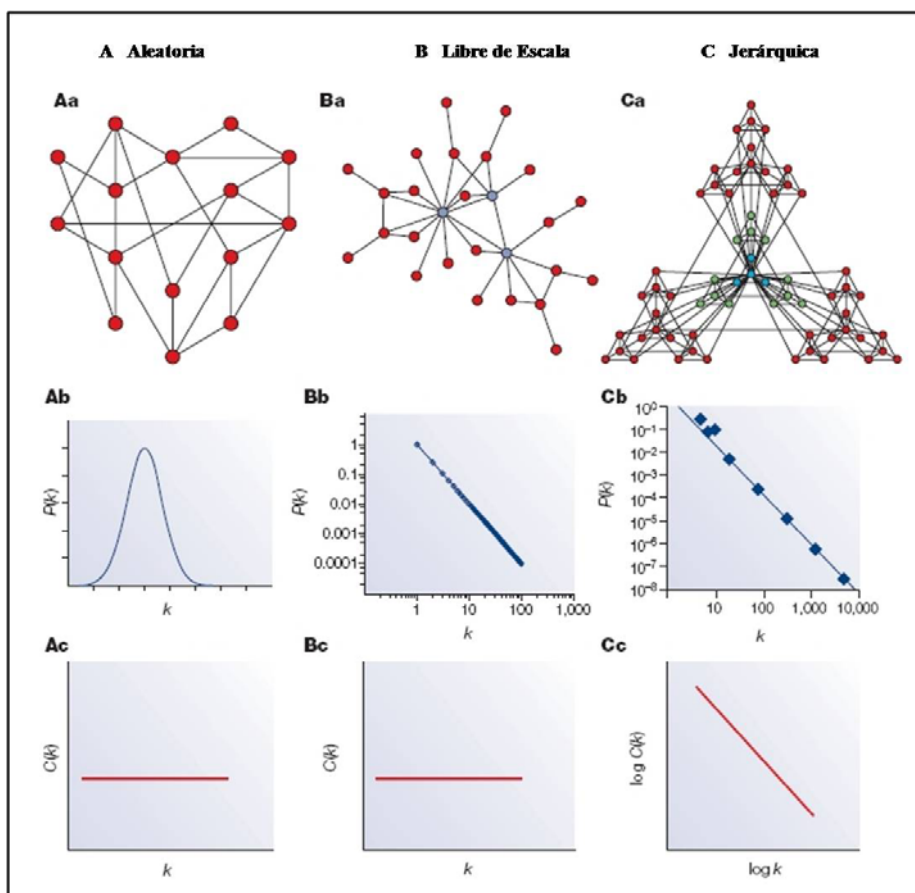


Figura 1. Modelos de redes y sus propiedades topológicas. A) Red aleatoria. Aa. Los nodos de ésta red se unen aleatoriamente. Ab. La distribución de la conectividad de los nodos sigue una distribución de Poisson. Ac. El coeficiente de clustering es independiente del grado de los nodos. B) Red libre de escala. Ba. En la red existen nodos altamente conectados (hubs), representados en azul. Bb. La distribución de las conectividades de los nodos sigue una ley de potencias. Bc. El coeficiente de clustering es independiente al número de conexiones de los nodos. C) Red Jerárquica. Ca. Los nodos se disponen en clústeres. Cb. La distribución de las conectividades de los nodos sigue una ley de potencias. Cc. El coeficiente de clustering sigue una ley de potencias. Figura obtenida del artículo (Barabasi & Oltvai, 2004).

1.3.1.1.1.1 Redes “libre de escala”.

Se ha descubierto que muchas redes biológicas presentan una estructura libre de escala. Esto implica una alta heterogeneidad en el número de conexiones de sus nodos. Mientras que la mayoría de los nodos de la red presenta una baja conectividad, unos pocos nodos tienen una alta conectividad, llamados *hubs*. (Barabasi & Bonabeau, 2003) (Figura 1Ba, 1Bb).

Una propiedad especialmente importante de las redes libres de escala, es la robustez de sus conexiones frente a la supresión aleatoria de nodos. En estas redes es más probable la eliminación de un nodo poco conectado que la de uno muy conectado, aunque la supresión de un *hub* puede producir un cambio en el sistema que cause su fragmentación (Albert *et al*, 2000).

Se ha explicado la existencia de *hubs* en la mayoría de las redes a partir de la combinación de dos mecanismos, crecimiento y acoplamiento preferente. Según el primero, la mayoría de las redes son el resultado de un proceso de crecimiento durante el cual nuevos nodos se añadirán al sistema tras un periodo de tiempo más o menos largo. Según el segundo, los nodos prefieren conectarse a nodos que ya tienen muchos enlaces (Barabasi & Albert, 1999).

1.3.1.1.1.2 Redes de “mundo pequeño”.

En 1967 el controvertido psicólogo social Stanley Milgram realizó un experimento bastante original: seleccionó cincuenta personas a las que entregó un mensaje para un único destinatario. El mensaje sólo podía ser entregado a un conocido y, por estos, a otro conocido, hasta alcanzar el objetivo final. Aunque el experimento no salió muy bien las primeras veces, en sucesivos intentos la tasa de recepción se elevó incluso hasta el 97%. Poco a poco emergió de los experimentos la idea de los “seis grados de separación”: cualquier persona podría llegar a cualquier otra siguiendo tan sólo seis pasos de “amigos de amigos” (Travers, 1969).

Duncan J. Watts y Steve Strogatz llegaron a la conclusión de que muchas redes biológicas presentan características similares, y en 1998 publicaron el célebre artículo en la revista Nature (Watts & Strogatz,

1998) donde proponen el modelo teórico que explica la existencia de estas redes de “mundo pequeño”. Este tipo de redes se encuentra entre una red regular y una aleatoria y se caracterizan por tener un *clustering* mayor que el de una red aleatoria y parecido al de una red regular, pero una longitud de caminos parecida al de una red aleatoria (Watts & Strogatz, 1998) (Figura 2).

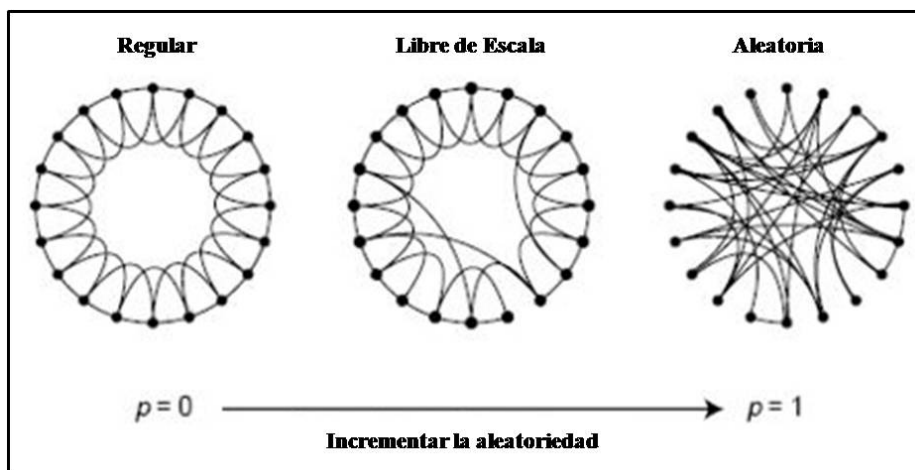


Figura 2. Recableado. La figura muestra el tipo de conectividad que presenta cada tipo de red. La red de libre de escala se encuentra en un punto intermedio entre una red regular y una aleatoria. Esta gráfica se ha obtenido del artículo (Watts & Strogatz, 1998).

Para el caso de grafos regulares, el diámetro crece linealmente con el número de nodos. Para grafos aleatorios, sin embargo, crece como el logaritmo del número de nodos. Este resultado expresa matemáticamente que es más rápido alcanzar cualquier punto desde un nodo escogido al azar en una red aleatoria que en una regular (Barabasi & Oltvai, 2004).

Lo que descubrieron Watts y Strogatz es que al introducir muy pocas conexiones al azar y probablemente a larga distancia, el diámetro de la red pasaba bruscamente de crecer linealmente a hacerlo logarítmicamente. La red adquiría esta propiedad típica de las redes aleatorias manteniendo el alto grado de clustering asociado a las redes regulares (Watts & Strogatz, 1998).

Se ha visto que las redes metabólicas son redes de “pequeño mundo”. En ellas los caminos entre nodos se corresponden con rutas metabólicas conectando dos substratos. Por tanto, las perturbaciones locales en la

concentración de un metabolito alcanzará muy rápidamente la red entera, recuperando la estabilidad del sistema rápidamente (Jeong *et al*, 2000; Wagner & Fell, 2001).

1.3.1.1.1.3 Jerarquía y modularidad en las redes.

El análisis de la modularidad en las redes biológicas ha sido también foco de estudio por su aportación en la comprensión de la estructura y comportamiento del sistema. El rasgo que define que un sistema sea jerárquico es que su coeficiente de *clustering* siga una ley de potencias ($C(k) \sim k^{-1}$), es decir, a medida que los nodos presentan una conectividad mayor, su coeficiente de *clustering* es menor (Figura 1C).

Cuando se habla de modularidad, se hace refería a un grupo de moléculas (nodos) físicamente o funcionalmente conectadas que trabajan juntas para realizar una función relativamente clara. Por lo tanto, en una red, un módulo o *cluster* aparece como un grupo de nodos muy interconectados (Hartwell *et al*, 1999; Alon, 2003; Ravasz & Barabasi, 2003).

En los sistemas biológicos, la modularidad jerarquizada concuerda con la noción de que la evolución puede actuar en varios niveles de organización simultáneamente. La acumulación de muchos cambios locales que afectan a módulos pequeños y altamente integrados, podrían lentamente tener impacto en las propiedades de módulos más grandes y menos integrados. De esta manera, la evolución podría operar copiando y rechazando módulos existentes para incrementar la complejidad del organismo. Bajo esta teoría, las redes jerárquicas surgirían a partir de la duplicación e integración de nodos que forman clústeres, proceso que en principio podría ser repetido indefinidamente (Barabasi & Oltvai, 2004). Por lo tanto, es comprensible la necesidad de conocer como esos módulos y la interacción entre ellos determinan la función del sistema. Existe un gran interés en explorar el origen de la evolución que contribuye a la formación de la red, y el impacto de esa estructura en funciones experimentalmente observables y características de su comportamiento (Parter *et al*, 2007; Tamames *et al*, 2007).

1.3.1.1.2 Topología de la red del Metabolismo Central.

Un sistema biológico que ha suscitado gran interés científico es el que describe los procesos metabólicos generales a muchos organismos, al cual se le denomina Metabolismo Central (CM) (Jeong *et al*, 2000; Teichmann *et al*, 2001b; Teichmann *et al*, 2001a; Wagner & Fell, 2001; Alves *et al*, 2002; Ravasz *et al*, 2002; Rison & Thornton, 2002; Schmidt *et al*, 2003). Gracias al esfuerzo realizado por grupos de investigación, como fue el caso del Centro de Bioinformática de la Universidad de Kyoto con el desarrollo de la base de datos KEGG (Kanehisa & Goto, 2000; Kanehisa *et al*, 2006), y a colaboraciones entre instituciones, es el caso del proyecto que dio lugar a la base de datos MetaCyc (Caspi *et al*, 2007), comenzaron a estar disponibles colecciones de datos con información muy completa de la mayoría de las rutas metabólicas acumulada durante décadas, dando paso a estudios sistémicos del metabolismo central.

MetaCyc es una base de datos de rutas metabólicas que ocurren en microorganismos y plantas, elucidadas experimentalmente. Alberga en su mayoría información cualitativa más que cuantitativa y su principal misión consiste en la predicción computacional de redes metabólicas en organismos que tengan su genoma anotado. Además puede utilizarse para diseñar alteraciones en el metabolismo de un organismo mediante ingeniería genética. Por su parte, KEGG (Kyoto Encyclopedia of Genes and Genomes) engloba una colección manual de rutas metabólicas que representan el conocimiento que de ellos se tiene a nivel de interacción molecular y la red de reacciones que forman. En ambos casos, las bases de datos engloban tanto información bioquímica (es decir, sus reacciones enzimáticas y rutas metabólicas) como genómica (genes encargados de realizar esos procesos metabólicos). Sin embargo, en KEGG normalmente se define una ruta metabólica combinando las pertenecientes a diferentes organismos, mientras que MetaCyc guarda de forma separada las rutas de cada organismo (Karp *et al*, 2002).

Esta información detallada de los procesos metabólicos permite estudios estructurales y topológicos de la red metabólica que forman. Gracias al análisis de las redes del Metabolismo Central de 43 organismos distintos, se observó que a pesar de las variaciones significativas en los compuestos y rutas metabólicas, todas ellas presentaban una topología libre de escala. Esto fue interpretado como que todas las redes de los organismos cumplían con unos principios de

diseño que aseguraban la robustez y tolerancia a errores del sistema (Jeong *et al*, 2000).

Por su parte Ravasz utilizando las redes del Metabolismo Central de esos 43 organismos, descubrió que todas ellas estaban organizadas en módulos muy pequeños y altamente conectados, los cuales se combinaban de manera jerárquica en unidades mayores y menos cohesionadas. Además se vio que en *Escherichia coli* ésta modularidad jerárquica prácticamente se solapaba con diferentes tipos de funciones metabólicas conocidas (Ravasz *et al*, 2002).

A su vez, Wagner y Fell (Wagner & Fell, 2001) utilizando la teoría de redes analizaron la red del Metabolismo Central de *Escherichia coli* encontrándose que ésta red era de “mundo pequeño”. A partir de este resultado propusieron que este tipo de estructura podía servir para minimizar los tiempos de transición entre estados metabólicos.

1.3.1.1.3 Antecedentes en el estudio de la topología de la red de Biodegradación.

Análogamente a lo que sucede en el metabolismo, numerosos estudios experimentales han aportado información acerca de la naturaleza bioquímica y genética de reacciones particulares y rutas específicas en microorganismos aislados (Warhurst *et al*, 1994; Seeger *et al*, 1995; Casellas *et al*, 1997). Utilizando toda esta información, en 1995, la Universidad de Minnesota desarrolló la base de datos UMBBD la cual recopilaba información acerca de numerosas rutas de biodegradación de compuestos químicos recalcitrantes realizadas por microorganismos (Ellis *et al*, 2006).

En 2005 Pazos realizó una base de datos llamada Metarouter a partir de la información disponible en la UMBBD. Esta base de datos fue diseñada para pronosticar el destino ambiental de compuestos recalcitrantes o mezclas de ellos, y de este modo definir estrategias en biorremediación (Pazos *et al*, 2005).

La biodegradación es comúnmente mediada por la coordinación de comunidades microbianas, las cuales transfieren sustratos y productos entre ellos, y contribuyen con los genes necesarios para llevar a cabo la degradación en las diferentes partes de la ruta. Esta cooperación metabólica es conocida desde hace mucho tiempo (Pelz *et al*, 1999;

Abraham *et al*, 2002). También se ha descrito que la transferencia horizontal de genes (HGT) es mucho más frecuente de lo que se había esperado (Wilkins, 2002) y que juega un papel clave a la hora de incorporar capacidades biodegradativas en comunidades bacterianas a través de la transferencia directa de genes catabólicos (Dejonghe *et al*, 2000). Por último se ha visto que los fenómenos atmosféricos movilizan y dispersan grandes cantidades de compuestos contaminantes a sitios lejanos de su lugar de origen (Carrera *et al*, 2002). Estas tres premisas hacen ver los procesos de biodegradación como una interconexión de reacciones (cooperación metabólica) donde los límites entre los organismos están difuminados (facilidad para incorporar nuevas capacidades por HGT) y sin una localización geográfica exacta (dispersión de los contaminantes).

Utilizando esta información, en 2003 Pazos realizó el primer estudio en biodegradación desde la aproximación de biología de sistemas. En él se analizó la red que formaban los compuestos que participan en los procesos de biodegradación, los compuestos inicial y final de las transformaciones. En este análisis se determinó que la red de compuestos de biodegradación poseía una estructura libre de escala al igual que la mayoría de las redes biológicas. También se observó que su comportamiento era similar al descrito en las redes metabólicas dentro de un solo organismo, lo que demostró la validez de considerar como un sistema al conjunto de reacciones de biodegradación independientemente del organismo (Pazos *et al*, 2003).

A su vez, en 2007 Gómez estudio la correlación entre la frecuencia de las triadas comunes en compuestos recalcitrantes y la capacidad de los microorganismos para metabolizarlos, desarrollando un sistema para predecir el destino metabólico de estos compuestos (Gomez *et al*, 2007).

Tanto la información contenida en la base de datos UMBBD y en la base de datos Metarouter, como los resultados obtenidos en el estudio realizado por Pazos en 2003 y Gomez en 2007 se orientan al análisis bioquímico de la red de biodegradación. El foco de atención en ambos casos se centra en las transformaciones de los compuestos más que en los genes que realizan esas transformaciones.

Sin embargo, como ya se ha señalado, se han realizado estudios experimentales donde se describe para ciertos organismos su potencial a la hora de participar en la transformación de compuestos

recalcitrantes y se ha identificado los complejos de proteínas encargados de dichas transformaciones (Warhurst *et al*, 1994; Seeger *et al*, 1995; Casellas *et al*, 1997). Haciendo uso de ésta información experimental y como complemento de los estudios sobre la red de compuestos de biodegradación descritos (Pazos *et al*, 2003; Pazos *et al*, 2005; Ellis *et al*, 2006; Gomez *et al*, 2007), parece necesario analizar la red de biodegradación fijándose en las enzimas que realizan las reacciones, más que en las transformación de los compuestos.

1.3.1.2 Evolución de las redes.

1.3.1.2.1 Teoría: Formación y evolución de las redes.

Todos los organismos vivos cuentan con rutas para sintetizar proteínas, ácidos nucleicos, lípidos y carbohidratos y catabolizar los compuestos que participan en los procesos celulares (Teichmann *et al*, 2001b). Se sabe que las enzimas son específicas y que catalizan reacciones particulares con sorprendente precisión y velocidad. Sin embargo para que se produzca evolución y adaptación a nuevos requisitos es necesario justamente lo opuesto: poder reconocer nuevos substratos y que las actividades enzimáticas evolucionen (Bork *et al*, 1993; Schmidt *et al*, 2003).

Aunque las rutas que realizan estos procesos podrían haberse creado “de novo”, se han propuesto varias teoría que explicarían la formación y evolución de estas redes (Lazcano & Miller, 1999; Schmidt *et al*, 2003) (Figura 3).

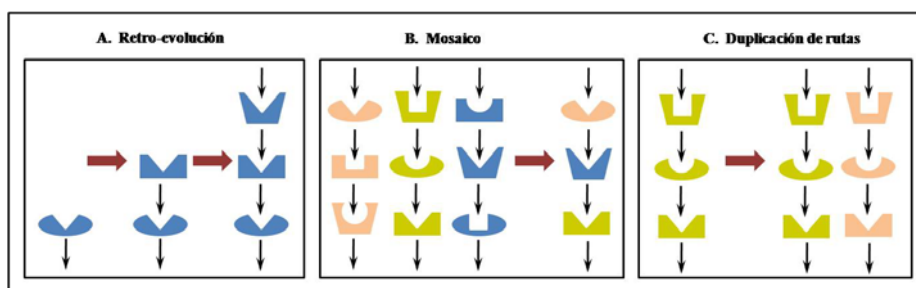


Figura 3. Teorías evolutivas en redes metabólicas. Los objetos representan enzimas, su forma indica la actividad catalítica que realizan y la forma de su muesca el tipo de substrato sobre el que actúan. A. En la teoría de retro-evolución los substratos de las enzimas son similares mientras que sus actividades catalíticas cambian a medida que la ruta se va formando. B. En el modelo de mosaico, la

formación de una nueva ruta se basa en reclutar enzimas de reacciones similares en diferentes rutas. C. En el modelo de duplicación de rutas, la nueva ruta se forma a partir de la duplicación de un conjunto de reacciones similares en otra ruta. La gráfica la he realizado basando en el artículo (Schmidt *et al*, 2003)

En 1945 Horowitz postulo que las rutas evolucionan “hacia atrás” desde un metabolito clave (Horowitz, 1945; Horowitz, 1965). A este modelo se le denominó “retro-evolución”. Un organismo usará un metabolito X hasta que este se agote. Si éste metabolito le era necesario para su crecimiento, supervivencia, etc, ser capaz de producirlo a partir de otro metabolito Y presente en el medio le dará una ventaja selectiva. Al utilizar este segundo metabolito Y llegará un momento que el metabolito Y también desaparezca, teniendo que ser capaz de obtenerlo a partir de otro precursor Z (repitiéndose el proceso). Horowitz afirma que la formación de las rutas metabólicas se basa en el mecanismo anterior y que por lo tanto al buscar precursores que vayan dando los compuestos que se van agotando, los metabolitos de la ruta tendrán una estructura química muy similar y por tanto las enzimas de esas rutas tendrán una unión a substrato parecida aunque una actividad catalítica diferente, cada paso metabólico requerirá una transformación diferente (Rison & Thornton, 2002; Schmidt *et al*, 2003). Sin embargo la aplicación del modelo de retro-evolución requiere que en el medio ambiente se encuentren acumulados compuestos orgánicos útiles y precursores potenciales. Aunque esta premisa podría haberse dado en el comienzo de la vida, estos compuestos deberían haberse convertido en menos comunes a medida que la vida se volvía más compleja y se agotaban en la naturaleza. Además, muchas rutas metabólicas producen compuestos intermedios que son inestables y podrían no persistir en la naturaleza el tiempo suficiente para provocar la presión selectiva necesaria que dé lugar al reclutamiento de una nueva enzima que catalice una nueva reacción.

En 1976, Jensen promulgó el modelo en mosaico (en inglés “patchwork”), por el cual las rutas habrían evolucionado desde enzimas ya existentes y poco específicas, las cuales se reutilizarían y combinarían para dar lugar a otras nuevas rutas. Esto produciría un mosaico de enzimas homologas a otras ya existentes dentro de la red que catalizan reacciones en rutas diferentes (Ycas, 1974; Jensen, 1976). Esta teoría se vuelve más probable conforme la vida se vuelve más compleja y el número de enzimas que pueden utilizarse incrementa considerablemente. Incluso este repertorio de actividades catalíticas disponibles se vuelve mucho mayor gracias a la capacidad de muchas enzimas para catalizar reacciones secundarias a un nivel, aunque bajo,

potencialmente útil. Esta promiscuidad catalítica de las enzimas puede ser utilizada para producir una nueva función cuando la realización de esa reacción secundaria aporte una ventaja selectiva al organismo.

Por último, existe un tercer modelo de evolución denominado “duplicación de rutas” el cual sugiere que bloques de varias reacciones consecutivas son duplicadas para divergir posteriormente y realizar una nueva función (Huynen & Snel, 2000; Rison & Thornton, 2002).

1.3.1.2.2 Evolución: Red del Metabolismo Central.

Numerosos estudios se han centrado en el análisis de la evolución del Metabolismo Central. Entre ellos, Sarah Teichmann, analizó los mecanismos que guiaban la formación de las rutas del Metabolismo Central de *Escherichia coli*. Los resultados obtenidos mostraron que por lo general las enzimas homólogas conservaban su actividad catalítica o unión a cofactor, mientras que su reconocimiento al sustrato en raras ocasiones se mantenía. También se vio que las enzimas homóloga se distribuían tanto a través de la red como dentro de la misma ruta y que el reclutamiento de dos enzimas consecutivas de una ruta a otra diferente era poco habitual. Por último observaron que la homología entre pares de enzimas consecutivas conservando la unión a sustrato pero cambiando su actividad catalítica ocurría de manera muy rara y muy puntual. Esto les llevo a proponer que el mecanismo principal en la formación de la red del Metabolismo Central era el modelo de mosaico, seguido del modelo de duplicación de rutas y muy por detrás el modelo de retro-evolución (Teichmann *et al*, 2001a; Teichmann *et al*, 2001b).

Éste estudio fue continuado por varios científicos. Por una parte Rison, corroboró los resultados anteriores y añadió un estudio filogenético cuyos resultados mostraron que la última enzima en una cadena metabólica no era necesariamente la más antigua. Ésta observación descartaba que la teoría de que la retro-evolución ha jugado un papel importante en el Metabolismo Central (Rison & Thornton, 2002). Por su parte, Schmidt definió superfamilias de enzimas como el conjunto de enzimas con origen común que podía ser detectado mediante homología de secuencia o estructural. Lo que observó es que mientras el 44% de las superfamilias realizaban solamente un tipo de actividad catalítica y sus enzimas participaban únicamente en un 10% de las reacciones de la red del Metabolismo Central, un 10% de las

superfamilias realizaban actividades catalíticas muy diferentes y sus enzimas participaban en el 45% de las reacciones del Metabolismo Central. Estos resultados muestran que los eventos de retro-evolución son muy poco habituales en el Metabolismo Central ya que únicamente un porcentaje pequeño de superfamilias son capaces de realizar actividades catalíticas diferentes (Schmidt *et al*, 2003).

Los estudios anteriores sobre la evolución, han focalizado su análisis al estudio del Metabolismo Central descrito para *Escherichia coli* y han condicionado su resultado a su clasificación en rutas metabólicas. En 2002, Alves y Sternberg enunciaron que limitar el estudio a un solo organismo y la definición de rutas producían un sesgo en los resultados obtenidos. Para solventar estos problemas utilizaron en su estudio los genomas de bacteria, totalmente secuenciados, disponibles hasta ese momento e introdujeron el concepto de distancia metabólica entre pares de enzimas de la red. Sus resultados mostraron que los pares de proteínas homologas solían estar separadas no más de tres pasos (evolución a nivel local de la red del Metabolismo Central) y que entre estas proteínas se solía conservar su actividad catalítica (Alves *et al*, 2002).

1.3.1.2.3 Antecedentes en el estudio de la evolución en la red de biodegradación.

Se ha analizado a nivel molecular los mecanismos por los cuales las bacterias han podido adaptarse rápidamente a la presencia de estos contaminantes en el medio (van der Meer *et al*, 1992; Top & Springael, 2003; Diaz, 2004; Springael & Top, 2004; Wackett, 2004).

En 1992 Van der Meer postuló que ésta adaptación genética está basada en la transferencia de genes, mutaciones puntuales y transposiciones y recombinaciones genéticas. Concluye su trabajo afirmando que para saber si estas enzimas especializadas en la degradación de compuestos recalcitrantes en las bacterias han evolucionado desde isoenzimas más comunes tras la introducción de grandes cantidades de productos en el medio ambiente, se necesitaría tener información acerca de la tasa de mutación en presencia de tales compuestos (van der Meer *et al*, 1992).

En 2003 Top insistió en la importancia de la transferencia horizontal de genes (HGT) en la adaptación de los microorganismos. Ésta HGT es

frecuentemente mediada por elementos genéticos móviles (MGEs) tales como transposones, integrones, islas genómicas, fágos y plásmidos fácilmente transferibles, como es el caso de los plásmidos IncP-2 y IncP-9, los cuales codifican la maquinaria de degradación de compuestos recalcitrantes naturales (Top & Springael, 2003). En 2004 Springael, continuó el estudio insistiendo en que la transferencia horizontal de genes y su ensamblaje en “mosaico” ocurren con una eficiencia relativamente alta y en un periodo de tiempo más o menos corto. También expuso la posible conexión entre las enzimas que degradan compuestos recalcitrantes y aquellas que transforman compuestos similares que aparecen en el medio ambiente de manera natural (Springael & Top, 2004).

También se ha estudiado la posible formación de ciertas rutas de biodegradación (Minshull, 1995; Copley, 2000; Janssen *et al*, 2005). En 2000 Copley describió como la ruta de degradación del pesticida pentaclorofenol en *Sphingomonas chlorophenolica* probablemente se había formado recientemente a partir del reclutamiento y posterior ensamblaje (modelo evolutivo en mosaico) de enzimas que provienen de la ruta de degradación de un fenol clorinado, que se encuentra de manera natural en el medio, y de la ruta metabólica de la tirosina. También señaló la ineficiencia del supuestamente reciente ensamblaje de las enzimas debido a una pobre eficiencia catalítica, a la inhibición de una enzima por su sustrato y a la carencia de regulación en la ruta, la expresión constitutiva de las enzimas asegura que van a estar disponibles cuando se necesiten pero implica un “derroche” (Copley, 2000).

1.3.2 Biología Sintética.

1.3.2.1 Contexto.

Al igual que la biología de sistemas, la biología sintética abre un abanico de posibilidades en el diseño e implementación de nuevas aplicaciones. En los años 60, se comenzó a descubrir la lógica matemática que se encontraba tras la regulación de los genes. Tal es el caso de Monod y Jacob, cuando en 1961 estudiaron el funcionamiento del operón “lac” (Jacob & Monod, 1961). Estos descubrimientos junto con la utilización de la tecnología de ADN recombinante a comienzos de los años 70 (Chakrabarty *et al*, 1973; Shaham *et al*, 1973), con la que es posible aislar y manipular un fragmento de ADN de un

organismo para introducirlo en otro, pusieron las bases de lo que más tarde se llamaría biología sintética.

La biología sintética es un nuevo campo de investigación que combina ciencia e ingeniería. Su objetivo es el diseño y la construcción de nuevas partes biológicas, dispositivos y sistemas y el rediseño de sistemas biológicos naturales para aplicaciones útiles. Su aplicación está condicionada al coste y al tiempo en el que se realice la síntesis del ADN, a la abstracción y la estandarización de la función biológica de las entidades, y al desarrollo de mejores herramientas para el estudio del comportamiento de los sistemas sintéticos (Tucker & Zilinskas, 2006).

Los esfuerzos que se están realizando en el campo se podrían clasificar en a) estandarizar y definir el comportamiento del mayor número de entidades biológicas (reguladores, activadores, represores, etc), de manera análoga a las hojas de especificación de los componentes electrónicos, para poder utilizarlas en la construcción de los sistemas biológicos (Endy, 2005; Bio *et al*, 2006), b) desarrollar e incorporar métodos y herramientas para el diseño y modelado de los circuitos biológicos (MathWorks, Inc.), c) rediseñar partes biológicas y mecanismos preexistentes para ampliar el juego de funciones que podemos usar y programar (Becskei & Serrano, 2000; Elowitz & Leibler, 2000; Gardner *et al*, 2000; Becskei *et al*, 2001), d) recrear y rediseñar nuevas funcionalidades en bacterias al ser organismos relativamente sencillos (Basu *et al*, 2005).

Al igual que en la electrónica se utilizan transistores, resistencias, condensadores, etc., para implementar circuitos, en la biología se utiliza material genético y el conocimiento de las interacciones (regulación) entre ellos para producir circuitos biológicos que simulen a los anteriores.

Siguiendo el paralelismo existente entre los circuitos electrónicos y la biología sintética, se han diseñado sistemas biológicos que responden a comportamientos descritos en los sistemas electrónicos que los modelan (Becskei & Serrano, 2000; Elowitz & Leibler, 2000; Gardner *et al*, 2000; Becskei *et al*, 2001). Este es el caso de los eventos de autorregulación. La retroalimentación es el proceso por el cual una parte de la señal de salida se vuelve a aplicar a la entrada. Esta retroalimentación es positiva si un aumento en la señal de salida da como resultado un aumento todavía mayor de la magnitud de la señal

de salida, actuando por lo tanto como un amplificador. Por el contrario esta retroalimentación será negativa si el sistema responde en dirección opuesta a la perturbación, haciendo que el nivel de salida del sistema se encuentre constante en un rango de funcionamiento.

En el caso biológico la autorregulación ocurre cuando una proteína modifica directa o indirectamente su propia tasa de producción. Serrano et al. diseñó tanto un sistema de retroalimentación negativa (Becskei & Serrano, 2000) como un sistema de retroalimentación positiva (Becskei *et al*, 2001), demostrando la estabilidad del primero de los sistemas frente a los sistemas no regulados o con retroalimentación positiva (Figura 4).

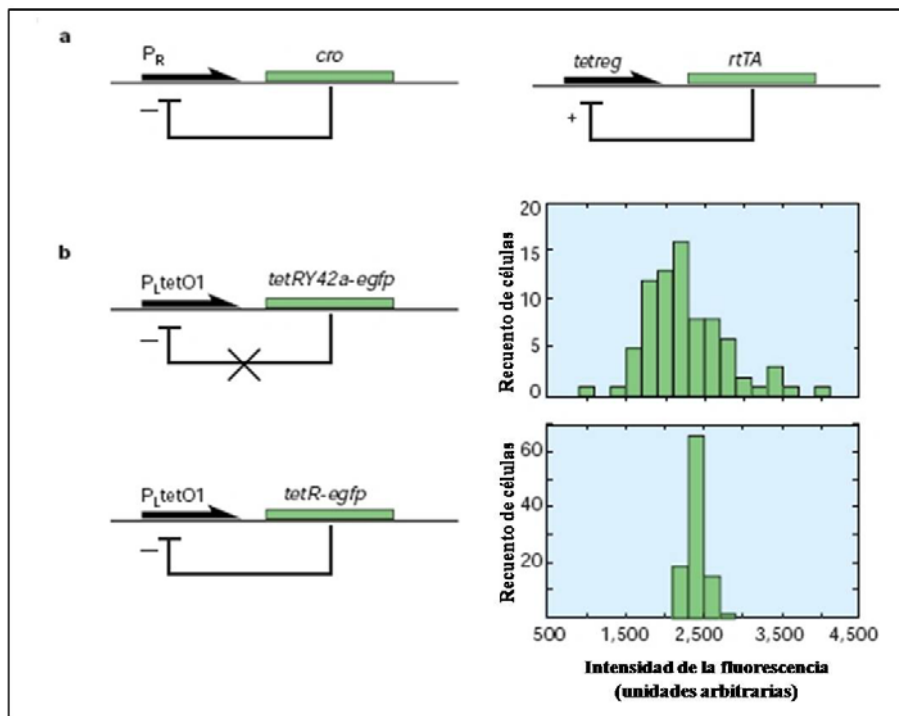


Figura 4. Sistemas de autoregulación. a) Este sistema natural de realimentación negativa en el bacteriofago, el promotor P_R controla la expresión de la proteína Cro la cual le reprime. b) Este es un sistema artificial de realimentación negativa usando el promotor P_{tetO1} para controlar la expresión de TetR-EGFP, una fusión del represor de tetraciclina (TetR) y una proteína florescente (EGFP). La realimentación negativa surge porque TetR reprime la transcripción de P_{tetO1} . Sustituyendo TetR por TetRY42A elimina la realimentación dando lugar a un sistema no regulado. En la figura se muestra como la distribución de la expresión para el sistema no regulado es tres veces más ancha que la distribución del sistema con realimentación negativa, demostrando el incremento de estabilidad que proporciona la realimentación negativa.

En otras ocasiones la realización de circuitos sintéticos ha ido dirigida a dotar a un organismo con nuevas capacidades. Este es el caso del trabajo realizado por Weiss et al. en el cual desarrolló en la levadura *Saccharomyces cerevisiae* un sistema de comunicación célula a célula utilizando para ellos elementos de señalización en *Arabidopsis thaliana* (Chen & Weiss, 2005).

1.3.2.2 Aplicación de la biología sintética en biodegradación.

Las expectativas puestas en la ingeniería genética como herramienta para la remediación “in situ” de problemas medioambientales, no se han visto cumplidas y aunque la biología sintética no es algo nuevo, ya que por ejemplo en 1987 el trabajo realizado por Rojo formaría parte de esta disciplina (Rojo *et al*, 1987), sí que se presenta como la evolución lógica y racional de la ingeniería genética en cuanto a sus aplicaciones medioambientales. De esta forma la aplicación de la biología sintética en biodegradación daría lugar al rediseño de organismos más eficientes en la descontaminación de ecosistemas.

Los esfuerzos en el campo de la biorremediación han sido dirigidos principalmente hacia la realización de biosensores de compuestos contaminantes. Los biosensores son dispositivos capaces de reconocer e interactuar con las sustancias de interés y de producir una señal de salida como consecuencia de esa interacción.

En algunos casos, la fabricación de estos biosensores consiste en la modificación del sitio de unión al ligando de una proteína para así reconocer el contaminante deseado. Este es el caso de la investigación realizada en la Universidad de Durham (Estados Unidos), en el cual se ha diseñado en *Escherichia coli* un biosensor capaz de detectar la presencia del explosivo Trinitrotolueno (TNT). La bacteria rediseñada es capaz de emitir una respuesta (en este caso fluorescencia) como consecuencia de su exposición al explosivo (Looger *et al*, 2003).

En otros casos, el biosensor se obtiene a partir de la incorporación de mecanismos que no son propios de las bacterias. Es el caso del biosensor realizado en *Saccharomyces cerevisiae* en donde se utilizó una ruta de señalización de mamíferos asociada con proteínas

fluorescentes, para reportar la presencia del explosivo 2,4-Dinitrotolueno (DNT) (Radhika *et al*, 2007).

Un paso más allá a la detección de contaminantes es la asociación al biosensor de un sistema de biorremediación. Para intentar biorremediar las contaminaciones por arsénico, en esta tesis se ha propuesto un sistema biológico capaz de eliminar el arsénico de lugares contaminados y por lo tanto actuar como “detoxificador”.

El arsénico es un caso muy representativo del daño que puede causar estos compuestos recalcitrantes al medio ambiente y a los seres humanos. El arsénico se encuentra de forma natural en la Tierra en pequeñas concentraciones y es extremadamente difícil de convertir en productos solubles en agua. Ha sido usado como insecticida (arseniato de plomo), herbicida (arsenito de sodio) y veneno, como pigmento y en pirotecnia (disulfuro de arsénico), como decolorante en la fabricación del vidrio (trióxido de arsénico) y como preservante de la madera (arseniato de cobre y cromo) (Jones, 2007).

Existen mecanismos muy diversos con los que los organismos eliminan el arsénico tras su absorción accidental. Uno de ellos es un sistema de bombeo que tienen bacterias como *Pseudomonas putida* o *Escherichia coli* gracias al cual tras la asimilación fortuita del arsénico, las bacterias son capaces de expulsarlo de nuevo al medio. Otro mecanismo es el que tiene el hígado del ratón y consiste en una proteína capaz de metilar el arsénico y así transformarlo en volátil y así expulsándolo al aire.

Para intentar biorremediar las contaminaciones por Arsénico, en esta tesis se ha desarrollado un sistema que combinando de manera artificial estos mecanismos naturales de defensa, es posible modular su actividad en función de la concentración de arsénico en el medio.

2 OBJETIVOS

2.1 Objetivo General.

Aplicar las metodologías propias de la Biología de Sistemas y la Biología Sintética al campo de la biodegradación, para determinar sus propiedades y origen, y su potencial en el desarrollo de aplicaciones.

2.2 Objetivos Específicos.

1. Desarrollar una base de datos de biodegradación que contenga información sobre rutas, reacciones, complejos, proteínas y genes y que su estructura permita el estudio de las relaciones entre estas entidades.
2. Realizar un estudio topológico de la red de reacciones de biodegradación.
3. Analizar la organización de las reacciones en la red de biodegradación.
4. Estudiar la formación y evolución de la red de reacciones de biodegradación.
5. Utilizando mi formación como ingeniero y los conocimientos adquiridos en el estudio de la red de biodegradación, diseñar, modelar y simular un circuito biológico: un sistema de detoxificación de arsénico.

3 MATERIALES Y MÉTODOS

3.1 Base de datos BioNeMo.

3.1.1 Asociación de cada nodo de la red (reacción) con su enzima.

Una metaenzima es el conjunto de complejos de proteínas que pueden realizar una reacción. La búsqueda de las metaenzimas se llevó a cabo partiendo de la información bioquímica que hay en la base de datos UMBBD (Ellis *et al*, 2006) y utilizando las base de datos de secuencias del NCBI (Maglott *et al*, 2007; Pruitt *et al*, 2007).

La información en la UMBBD que se utilizó fue el número EC y el nombre de la actividad catalítica de la reacción, el artículo donde se describe esa transformación bioquímica, el nombre del autor del artículo, el organismo y el nombre de la ruta a la que pertenece. Sin embargo, a medida que iba recopilando información de las reacciones de una ruta, también la utilizaba para obtener información de otras reacciones, por ejemplo viendo si en el mismo operón existían genes que codificaban proteínas que realizaban esa transformación.

Las búsquedas las realicé de manera manual debido a que la combinación de datos que da un resultado positivo es para cada reacción totalmente diferente. Introduciendo esas combinaciones en el motor de búsqueda “Entrez” del NCBI (Maglott *et al*, 2007) y supervisando los resultados manualmente, asocié los complejos a las reacciones.

La agrupación de las proteínas asociadas a una reacción en complejos enzimáticos, se realizó utilizando la información que tiene las entradas de GenBank de cada proteína donde se indica el complejo enzimático al que pertenecen.

3.1.2 Realización de la base de datos.

La base de datos BioNeMo ha sido implementada utilizando PostgreSQL (*Berkeley Software Distribution*). Las tablas más significativas desde el punto de vista enzimático son la de rutas, reacciones, complejos, proteínas, secuencia de proteínas, homología,

pfam y pdb. Estas y sus relaciones aparecen en la siguiente gráfica (Figura 5).

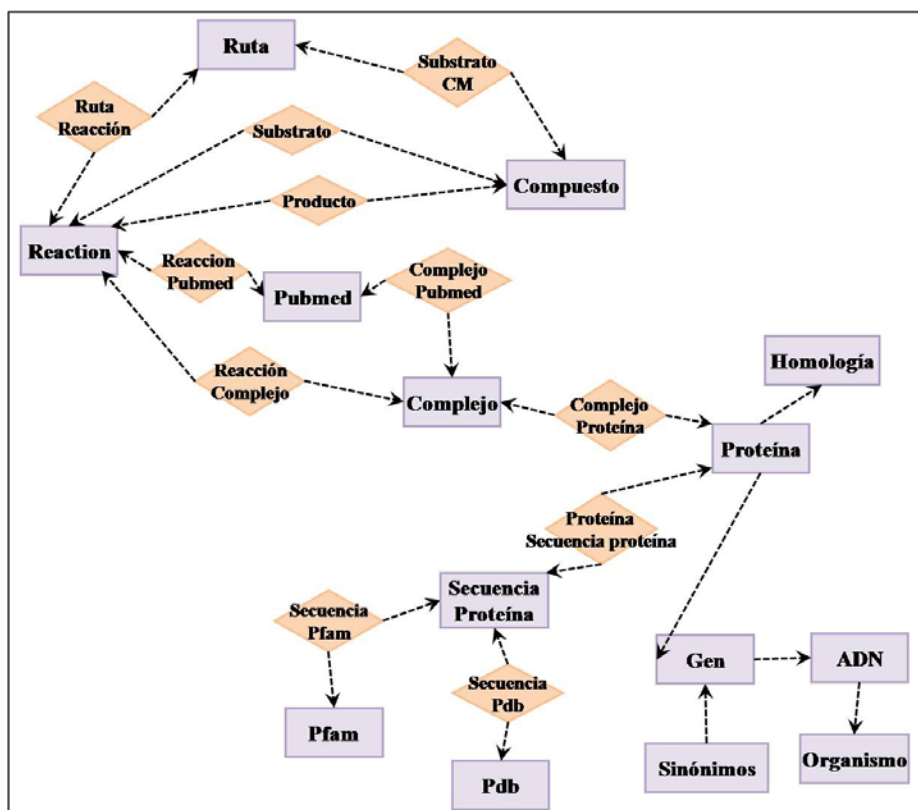


Figura 5. Estructura de la información "enzimática" en la base de datos BioNeMo. Los cuadrados morados representan tablas de entidades, mientras que los rombos azules tablas de conexión.

El contenido de las tablas más importantes de la parte enzimática de BioNeMo se describe a continuación.

Tabla ruta		
Campo	Tipo	Ejemplo
id_pathway	Entero	10
name	Carácter	Benzoate
abrev	Carácter	benz
tipo	Carácter	anaeróbica

Tabla ruta_reacción		
Campo	Tipo	Ejemplo
id_pathway	Entero	9
id_reaction	Entero	218

Tabla reacción		
Campo	Tipo	Ejemplo
id_reaction	Entero	218
minnesota_code_react	Carácter	r0245

Tabla reacción_complejo		
Campo	Tipo	Ejemplo
id_complex	Entero	657
id_reaction	Entero	218
ec_code	Carácter	1.14.12.10

Tabla complejo		
Campo	Tipo	Ejemplo
id_complex	Entero	657
activity	Carácter	ENZ
description	Text	benzoate 1,2-dioxygenase

La actividad de los complejos puede ser enzimático (ENZ) o regulación (REG), aunque en esta tesis solo considero los complejos enzimáticos.

Tabla complejo_proteína		
Campo	Tipo	Ejemplo
id_complex	Entero	657

id_protein	Entero	828
------------	--------	-----

Tabla proteína		
Campo	Tipo	Ejemplo
id_protein	Entero	828
gb_protein	Carácter	CAG68302
id_gene	Entero	828
code_sw	Carácter	sp: BENC_ACIAD

Tabla proteína_secuencia_proteína		
Campo	Tipo	Ejemplo
id_sequenc	Entero	828
id_protein	Entero	828
gi_number	Entero	-

Tabla secuencia_proteína		
Campo	Tipo	Ejemplo
id_sequenc	Entero	828
Sequence	Carácter	MSLYLNRIPAMSNHQVALQFED GVTRFIRIAQGETLSDAAYRQQ INIPMDCREGACGTCR AFCESGNYDMPEDNYIEDALTP EEAQGGYVLACQCRPTSDAVFQ IQASSEVCKTKIHHFE GTLARVENLSDSTITFDIQLDD GQPDIIHFLAGQYVNVTLPGTTE TRSYSFSSQPGNRLTG FVVRNVPQGKMSEYLSVQAKAG DKMSFTGPFGSFYLRDVKRPVL MLAGGTGIAPFLSMLQ VLEQKGSEHPVRLVFGVTQDCD LVALEQLDALQQKLPWFYRTV VAHAESQHERKGYVTG HIEYDWLNGGEVDVYLCGPVPM VEAVRSWLDTQGIQPANFLFEK

	FSAN
--	------

En las tablas “secuencia_pfam” y “pfam” se guardan los resultados de aplicar modelos ocultos de markov obtenidos de Pfam a las secuencias de las proteínas que participan en procesos de biodegradación contra la base de datos de dominios Pfam (Finn *et al*, 2007).

Tabla secuencia_pfam		
Campo	Tipo	Ejemplo
id_sequence_pfam	Entero	1090
id_sequence	Entero	828
id_pfam	Entero	96
query_start	Entero	19
query_end	Entero	99
hmm_start	Entero	1
hmm_end	Entero	95
bit_score	Carácter	89,1
evaluate	Carácter	1,3e-23

Tabla pfam		
Campo	Tipo	Ejemplo
id_pfam	Entero	96
code_pfam	Carácter	PF00111.17
description	Texto	2Fe-2S iron-sulfur cluster binding domain
ga_ls_sequence	Carácter	7.0
ga_ls_domain	Carácter	7.0
tc_ls_sequence	Carácter	7.0
tc_ls_domain	Carácter	7.0
nc_ls_sequence	Carácter	6.9
nc_ls_domain	Carácter	6.9

En la tablas “secuencia_pdb” y “pdb” se guardan los resultados del Blast de las secuencias de las proteínas que participan en reacciones de biodegradación contra la base de datos de Pdb (Berman *et al*, 2000; Berman *et al*, 2002), con un “e-value” de $1 \cdot 10^{-6}$.

Tabla secuencia_pdb		
Campo	Tipo	Ejemplo
id_sequence_pdb	Entero	1254
id_sequence	Entero	828
id_pdb	Entero	398
query_start	Entero	11
query_end	Entero	348
hit_start	Entero	1
hit_end	Entero	338
long_query	Entero	348
long_hit	Entero	338
long_align	Entero	338
total_coverage	Carácter	N

Tabla pdb		
Campo	Tipo	Ejemplo
id_pdb	Entero	398
code_pdb	Carácter	1KRH
chain	Carácter	B

En la tabla “homología” se guardan los resultados del Blast de cada una de las proteínas que participan en reacciones de biodegradación contra la base de datos Swissprot y Trembl (versión Enero 2007).

Tabla Homología		
Campo	Tipo	Ejemplo
id_homology	Entero	40644
id_query	Entero	828
sw_code_query	Carácter	BENC_ACIAD
query_start	Entero	14
query_end	Entero	348
query_length	Entero	348
id_hit	Entero	1333
sw_code_hit	Carácter	Q0WZ45_PSEPU
hit_start	Entero	3
hit_end	Entero	335
hit_length	Entero	336
score	Carácter	381
evaluate	Carácter	2e-107
alignment_length	Entero	335
percentage_identities	Doble Precisión	55,22

Una vez que había asociado y almacenado en BioNeMo para cada reacción sus complejos enzimáticos, y para cada complejo sus proteínas, utilizando sus secuencias obtenidas de GenBank, anoté el código Uniprot de las proteínas a través de un Blast contra Swissprot y Trembl con un “e-value” de $1 \cdot 10^{-6}$. Si una secuencia de proteína de BioNeMo tenía en Swissprot una con el 100% de porcentaje de identidad, una longitud que no se diferenciara en más de un 10% y que su alineamiento al menos la cubriera en un 90%, entonces le asignaba ese código Swissprot. En caso contrario sí se encontraba en Trembl una secuencia que cumpliera las condiciones anteriores, entonces asociaba el código Trembl a esa proteína de BioNeMo.

3.1.3 Realización del servidor web.

La estructura que he seguido para desarrollar el servidor web de la base de datos BioNeMo es el siguiente. Parto de una interfaz de entrada común ("Common Gateway Interface", CGI) que permite a un explorador web solicitar datos de un programa ejecutado en un servidor web. Este programa usa y llama un modulo de aplicación que ejecuta y maneja los modos de ejecución de la aplicación. Cada modo de ejecución se corresponde con una página web (la página de búsqueda/inicio, la de resultados "google", la de reacciones, la de genes/proteínas, la de rutas y la de unidades de transcripción). Cada uno de estos modos de ejecución contiene funciones que se encargan de recopilar la información necesaria en cada página. Una vez que cada modo ha recopilado la información, ejecuta un módulo "HTML::Template" que contiene el código HTML de cada una de las páginas del servidor web de BioNeMo. Dependiendo de los parámetros que se le pasen una parte del programa será ejecutada (aquella que se corresponda con el código HTML de la página a mostrar). Esta estructura permite independizar el código Perl (que se muestra) del código HTML (como se muestra).

3.2 *Análisis del nivel de descripción de la red de Biodegradación.*

3.2.1 Formación de la red de biodegradación.

En esta tesis se ha optado por representar los procesos de biodegradación de compuestos recalcitrantes mediante una aproximación centrada en las reacciones en lugar de la clásica representación de los procesos metabólicos centrados en la transformación de los compuestos. Por tanto, la red de reacciones de biodegradación es una red dirigida en donde los nodos son las reacciones y los enlaces son los compuestos (Figura 6).

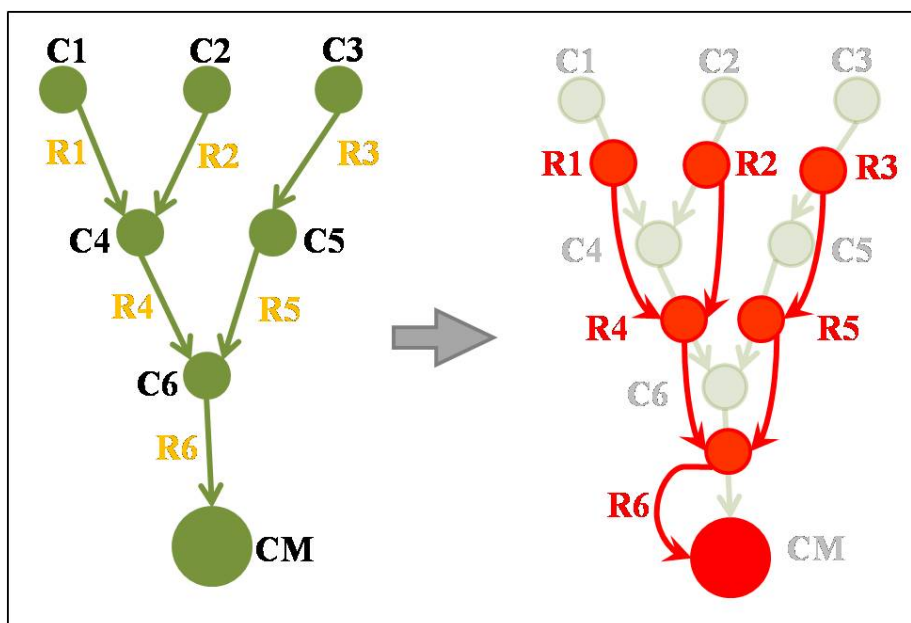


Figura 6. Transformación: red de compuestos a red de reacciones. La red de la izquierda representa una red de compuestos en la cual los nodos (círculos) son compuestos y los enlaces (flechas) las reacciones de las que son sustratos y productos. Dos compuestos están unidos (por ejemplo C3 y C5) si hay una reacción que transforma el primero en el segundo (en este caso "R3"). Este tipo de red es la estudiada por Pazos (Pazos et al, 2003) y la descrita en las bases de datos de UMBBD y Metarouter. Su transformación a una red de reacciones se muestra en la figura de la derecha. En este caso dos reacciones (R1 y R4) están unidas si un producto de la primera es sustrato de la segunda (en el ejemplo C4). Ambas redes llegan a un nodo que representa la entrada al Metabolismo Central (CM).

Dos nodos o reacciones, A y B, están conectados con una determinada dirección cuando alguno de los productos de la reacción A se corresponde con alguno de los substratos de la reacción B. Cuando una reacción tiene más de un substrato o producto, he construido todas las posibles conexiones. Los compuestos químicos anotados como cofactores en UMBBD no se han considerado para construir la red. Todas las reacciones en UMBBD se han incluido en la red independientemente de su naturaleza aeróbica o anaeróbica y el organismo en el cuál se ha descrito que se realiza. La red se compone de 945 reacciones.

La definición de las rutas de biodegradación en las que participan las reacciones también se ha tomado de la base de datos de UMBBD. Al mismo tiempo, debido a que muchas de estas rutas de biodegradación producen un compuesto final que pertenece al metabolismo central, se ha introducido en la red un nodo artificial que representa la entrada a la red del Metabolismo Central.

3.2.2 Comparación de la red de Biodegradación con la red del Metabolismo Central.

La red del Metabolismo Central se ha obtenido a partir de las rutas del metabolismo común definidas en KEGG (Kanehisa & Goto, 2000; Kanehisa *et al*, 2006). Las rutas utilizadas se describen en el Apéndice 1.

El nivel de descripción de las reacciones del Metabolismo Central se obtuvo de la siguiente manera. Por una parte se identificó las reacciones que participan en estas rutas utilizando el fichero “reaction” que suministra KEGG en la dirección <ftp://ftp.genome.jp/pub/kegg/ligand/reaction/reaction>. Solo se tomaron las reacciones para las que los dígitos de la ruta de la que forma parte (rnXXXXX) coincidiera con las de la lista anterior (mapXXXXX). De este fichero también se obtuvieron el código de ortología utilizado por KEGG (KO) el cuál es un sistema de clasificación de genes ortólogos incluyendo grupos de genes parálogos. KEGG utiliza este código para asignar a las reacciones, los genes que las realizan.

Una vez que cada reacción tiene asignado su código KO, utilizando el fichero ftp://ftp.genome.jp/pub/kegg/linkdb/genes/genes_ko.list asocio

a cada grupo de ortólogos (KO) sus genes, para posteriormente utilizando el fichero ftp://ftp.genome.jp/pub/kegg/linkdb/genes/genes_uniprot.list identificar el código Uniprot que tiene cada uno de los genes. Relacionando estas informaciones, puedo saber si la reacción tiene una proteína en Swissprot/Trembl asociada o no.

Para saber si las reacciones del Metabolismo Central tienen asociado dominios funcionales (Pfam) y estructuras tridimensionales (PDB), ya que tengo los códigos Uniprot (Swissprot o Trembl) he buscado la información en esa base de datos (2007).

3.2.3 Comparación de las reacciones que llegan al Metabolismo Central de las que no.

Considerando que la entrada de los compuestos en el Metabolismo Central es un nodo más en la red de reacciones de biodegradación, una reacción se conecta directamente con éste nodo cuando alguno de sus productos está asignado en UMBBD como que forma parte de la red del Metabolismo Central. De esta manera, una reacción se considera que forma parte de un proceso de biodegradación completo cuando forma parte de una ruta en la que alguno de los productos producidos entran al Metabolismo Central, es decir, podemos conectar (a través de una serie de reacciones) a esta reacción con una reacción en donde alguno de sus productos conectan con el nodo del Metabolismo Central. Por el contrario, si la reacción no puede conectarse con una reacción que enlace con el Metabolismo Central, esa reacción formará parte de una biodegradación parcial.

3.2.4 Distancia al Metabolismo Central.

La distancia desde un determinado nodo (reacción) al Metabolismo Central se define como el mínimo número de enlaces que les separan en la red. Para calcularlo he utilizado el algoritmo de Dijkstra y programación recursiva. La idea subyacente en este algoritmo consiste en ir explorando todos los caminos más cortos que parten del vértice origen y que llevan a todos los demás vértices; cuando se obtiene el camino más corto desde el vértice origen, al resto de vértices que componen el grafo, el algoritmo se detiene.

3.3 Topología de la red de Biodegradación.

Para el estudio de la topología de la red de reacciones de Biodegradación utilicé las relaciones entre las reacciones contenidas en la base de datos de la Universidad de Minnesota (UMBBD). La metodología utilizada para ver si cada una de estas características se cumplen en la red de biodegradación, la he obtenido de la revisión sobre redes biológicas publicada por Barabási, en la cual se describe con detalle los pasos a seguir (Barabasi & Oltvai, 2004).

3.3.1 Red libre de escala.

Para averiguar si la estructura de la red es libre de escala, he calculado la probabilidad de que una enzima tenga un cierto número de conexiones. Una red se dice que es libre de escala, cuando la mayoría de los nodos tienen muy pocas conexiones y solo un número pequeño de ellos están muy conectados (llamados *hubs*). Por lo tanto en una red de este tipo la probabilidad de que un nodo tenga un número de enlaces (k) sigue una ley de potencias ($p(k) \sim k^{-g}$) donde “ g ” es el grado del exponente.

3.3.2 Red “mundo pequeño”.

Una red de “mundo pequeño” se caracteriza por tener un coeficiente de *clustering* mayor que el de una red aleatoria pero una longitud de caminos parecida al de una aleatoria. El coeficiente de *clustering* de un nodo “ I ” se define por la ecuación siguiente: $C_I = 2n_I / (k(k-1))$ donde n_I es el número de enlaces conectando el nodo I con sus k_I vecinos, y $k(k-1)$ es el número máximo de conexiones que se podrían establecer si todos los vecinos del nodo I estuvieran conectados unos con otros. A su vez para calcular la longitud de caminos entre los nodos de la red he utilizado el algoritmo de Dijkstra.

La red aleatoria que he utilizado para realizar las comparaciones la he construido con el programa Matlab (MathWorks, Inc.). Ésta consta del mismo número de nodos y conexiones totales que la red de reacciones de biodegradación, y donde cada nodo además mantiene el número de conexiones de entrada y de salida. Se realizó utilizando el programa

Matlab (MathWorks, Inc.) para a partir de la red de reacciones de biodegradación obtenerla. He utilizado Perl (*Practical Extraction and Report Language*) para obtener los resultados.

3.3.3 Red jerárquica.

En la red del Metabolismo Central se ha observado que los clústeres (conjunto de nodos muy conectados) se combinan generando una estructura jerarquizada. Para esto, el coeficiente de *clustering* de la red ($C(k)$) debe disminuir a medida que la conectividad (k) aumenta siguiendo una ley de potencias ($C(k) \sim k^{-1}$), es decir cuanto mayor sea la conectividad de un nodo, sus nodos vecinos estarán menos conectados. He utilizado Matlab (MathWorks, Inc.) para obtener los resultados.

3.4 Organización de la red de Biodegradación.

3.4.1 Complejidad.

Los datos utilizados en el estudio de la complejidad de las reacciones se han obtenido utilizando la base de datos BioNeMo. De esta manera el número de dominios funcionales que tiene una reacción se extrae viendo para cada reacción de biodegradación, los complejos enzimáticos que puedan realizarla, las proteínas que forman parte de estos complejos y los dominios funcionales asociados a las secuencias de estas proteínas. Esta compleja consulta a la base de datos, se realiza fácilmente haciendo uso de la API.

Para conocer el número de dominios funcionales de cada proteína y el número de proteínas por complejo también he utilizado la base de datos BioNeMo. En este caso las consultas son muy sencillas y directas utilizando la API (proteína -> pfam; complejo enzimático -> proteínas).

La distribución de los datos anteriores (número de dominios por reacción, número de proteínas por complejo, número de dominios funcionales por proteína) en la red de reacciones de biodegradación se obtiene al introducir la distancia entre la reacción que tiene esos dominios, para el primero de los datos anteriores, o en la que participa cada complejo, para el segundo de los datos, o de la que forma parte las proteínas, para el tercero de los datos, al nodo que representa la entrada al Metabolismo Central (ver apartado 3.2.4 para detalles del método).

De esta forma se obtiene el porcentaje de reacciones en cada distancia que tiene un número de dominios funcionales, el porcentaje de complejos en cada distancia que tienen un número de proteínas, teniendo en cuenta que si un complejo aparece en más de una reacción, se contabiliza tantas veces como reacciones diferentes participe y en las distancias al Metabolismo Central en las que esas reacciones se sitúen, y el porcentaje de proteínas en cada distancia que tienen un número de dominios funcionales, si una proteína forma parte de un complejo que participa en más de una reacción, aparecerá en todas las distancias de esas reacciones y tantas veces como reacciones esté asociada.

3.4.2 Especificidad.

La especificidad de los complejos hace referencia al número de reacciones que un mismo complejo puede llevar a cabo. Para analizar la distribución de la especificidad de las reacciones dentro de la red de biodegradación realicé varios pasos. Primero identifiqué para cada complejo de proteínas el número de reacciones de biodegradación que realizaba, haciendo uso de la base de datos BioNeMo y de su API. Segundo, calculé la distancia a la que se encontraba cada uno de los complejos viendo la distancia de las reacciones que podía realizar. De esta forma, un mismo complejo se contabiliza tantas veces como reacciones participe y en las distancias en las que se encuentren estas reacciones. Por tanto, los datos obtenidos muestran en cada distancia la especificidad de todos los complejos de las reacciones a esa distancia.

3.4.3 Variabilidad.

Para calcular la variabilidad de las proteínas de biodegradación, he utilizado las proteínas homologas a las de biodegradación almacenadas en la base de datos BioNeMo. Esta colección de proteínas se ha obtenido mediante un Blast de las secuencias de las proteínas en BioNeMo contra las bases de datos Swissprot y Trembl con “e-value” de $1 \cdot 10^{-6}$. De todos los resultados obtenidos solo se han almacenado en BioNeMo cuando su diferencia de longitudes no era superior a un 10% y el alineamiento no difería más de un 10% de la longitud de la secuencia de la proteína de biodegradación.

A partir de esta colección y fijando en cada caso un porcentaje de identidad mínimo diferente (95% y 50%) se ha estudiado como afecta el porcentaje de conservación de la secuencia en el número de homólogos de las proteínas.

La distribución de la redundancia a lo largo de la red de biodegradación se ha calculado al asignar a cada proteína que forma parte de un complejo enzimático, la distancia o distancias de las reacciones en las que participa (si una proteína forma parte de un complejo que realiza varias reacciones, la proteína aparecerá en el estudio tantas veces como reacciones participe).

3.5 Evolución de la red de Biodegradación.

3.5.1 Dominios funcionales de la red de Biodegradación y la del Metabolismo Central.

Para cada dominio funcional asociado a proteínas que forman parte de alguno de los complejos enzimáticos de la red de biodegradación y que se encuentran almacenados en la base de datos BioNeMo, he calculado el porcentaje de reacciones de biodegradación en los que aparece.

Utilizando el mismo grupo de dominios funcionales y la asociación antes definida entre las reacciones del Metabolismo Central y sus dominios funcionales, he calculado el porcentaje de reacciones en el Metabolismo Central donde los dominios funcionales de biodegradación están presentes. Relacionando estos dos porcentajes he realizado la comparativa funcional entre las dos redes.

3.5.2 Evolución de las proteínas.

La relación evolutiva de las proteínas de la red de biodegradación con la de la red del Metabolismo Central se obtuvo primero realizando un blast a partir de las proteínas que estaban asociados a un complejo enzimático en biodegradación y que están almacenadas en la base de Datos Bionemo, contra la colección de secuencias de proteínas de Swissprot y Trembl. Segundo, como para cada reacción del Metabolismo Central (las que forman parte de las rutas de KEGG seleccionadas), tengo el código Uniprot que tienen sus proteínas (ver apartado 3.2.2 para detalles del método), comprobé si en los resultados del Blast existían proteínas del Metabolismo Central.

Dos proteínas se consideran homólogas si el porcentaje de identidad entre sus secuencias es mayor del 30%, para estar segura de que existe relación evolutiva (Rison & Thornton, 2002), la diferencia de longitudes entre ellas es menor del 10% y la diferencia entre la longitud del alineamiento y las longitudes de las secuencias no es superior al 10%.

3.5.3 Evolución de las reacciones.

La relación evolutiva entre las reacciones de Biodegradación y del Metabolismo Central se calcula a partir de la relación evolutiva de las proteínas que forman parte de sus complejos enzimáticos (ver apartado 3.5.2 para más detalles). Los pasos a seguir son: primero, identificar la conexión entre las proteínas, los complejos enzimáticos a los que pertenecen y las reacciones de biodegradación en las que participan utilizando para ello la base de datos BioNeMo. Segundo, definir para cada reacción, con al menos un complejo asociado, el “complejo más informativo”, el cual se corresponde con el complejo que tiene un mayor número de subunidades (proteínas). Esto podemos realizarlo al haber visto que excepto para dos reacciones, los complejos de proteínas asociados a una reacción tienen proteínas homólogas. Tercero, identificar la relación evolutiva de las proteínas de los “complejos más informativos” mediante comparación de secuencia. De esta forma, si todas las proteínas del complejo tienen una relación evolutiva detectable con el Metabolismo Central, clasifico la reacción como “CM”, si ninguna de las proteínas del complejo lo tienen, clasifico la reacción como “no CM” y si no se cumplen los casos anteriores clasifico la reacción como mixta.

Para analizar la robustez de los resultados, he repetido el experimento utilizando mil redes aleatorias análogas a la de biodegradación (es decir, con el mismo número de reacciones, y donde cada complejo asociado a la reacción tiene el mismo número de proteínas) pero con la diferencia de que las proteínas que forman cada uno de esos complejos es aleatoria.

La manera en la que la he realizado es la siguiente. Tomando al azar dos reacciones de la red que tengan “complejo más informativo” asociado, selecciono al azar de cada uno de los complejos una proteína y intercambiándoselas. Este proceso lo repito 1400 veces y al terminar de realizarlo cuento el número de reacciones “CM”, “no CM” y “mixtas” que hay en esa red y los almaceno guardando. Todo esto se repetirá a su vez 1000 veces, pudiendo hacer una distribución del número de reacciones de cada tipo con la colección de datos almacenada.

El valor del z-score de un evento indica como de lejos y en qué dirección este evento se desvía respecto a la media de una distribución, expresándose en fracciones o múltiplos de la desviación estándar de esa

distribución. Se calcula mediante la fórmula siguiente: $Z = (x - \mu_x)/\sigma_x$, donde x es el valor para el cuál quiero calcular el z-score, μ_x la media de la distribución con la que quiero comparar el valor y σ_x la desviación de esa distribución.

3.5.4 Bloques de reacciones consecutivas.

Para analizar si dos reacciones consecutivas de la red de biodegradación que tuvieran relación evolutiva con el Metabolismo Central, sus reacciones relacionadas por similitud de secuencia en el Metabolismo Central también se encontraban consecutivas (seguidas y manteniendo el orden de aparición, ya que son redes direccionales) es necesario saber además de las proteínas asociadas a cada reacción del Metabolismo Central (ver apartado 3.2.2), las conexiones entre las reacciones. Esto implica formar la red de reacciones del Metabolismo Central. Para ello utilizando el fichero facilitado por KEGG <ftp://ftp.genome.jp/pub/kegg/ligand/reaction/reaction>, obtengo los sustratos y productos de cada una de las reacciones. Con ellos, dos reacciones A y B están conectadas con un sentido (A->B) si alguno de los productos de A son sustrato de B. Si una reacción tiene varios sustratos o productos se establecen conexiones a partir de todos sus sustratos y productos. Una vez formada la red de reacciones del Metabolismo Central, para cada par de reacciones de biodegradación consecutivas y con relación evolutiva en el Metabolismo Central, se identificó exactamente con que reacciones del Metabolismo Central se relacionaban y si están se encontraban conectadas consecutivamente. Si la búsqueda era positiva, entonces el par de reacciones de biodegradación tienen una relación evolutiva con el Metabolismo Central y han mantenido su ensamblaje. En caso contrario, las reacciones solo tienen relación evolutiva con el Metabolismo Central.

Igual que en el apartado anterior he repetido mil veces el experimento anterior utilizando redes aleatorias análogas a la de biodegradación. En este caso se trata de redes donde se mantienen las conexiones de la red de biodegradación y el número de reacciones de cada tipo (CM, no CM y mixtas).

La manera en la que la he realizado es la siguiente. Tomando al azar dos reacciones de la red que tengan “complejo más informativo” asociado, intercambio sus complejos. Este proceso lo repito 1400 veces

y al terminar de realizarlo cuento el número de reacciones que siendo consecutivas en la red y siendo las dos del tipo CM, los complejos homólogos a ellos en el Metabolismo Central también estén consecutivos o no y almaceno guardando. Todo esto se repetirá a su vez 1000 veces, pudiendo hacer una distribución del número de bloques CM con el mismo ensamblaje o diferente ensamblaje que en el Metabolismo Central con la colección de datos almacenada

3.5.5 Teorías de evolución de las redes metabólicas.

Para este análisis he utilizado el definido el concepto de par de reacciones relacionadas por similitud de secuencia (RRSS). Dos reacciones de biodegradación podrán formar un par de RRSS si las proteínas de sus complejos más informativos (ver apartado 3.5.4) son homólogos. Si los complejos tienen un número diferente de proteínas, son homólogos si todas las proteínas del complejo menor son homólogos a alguna de las proteínas del complejo mayor. La homología entre las proteínas de ambos complejos se define de la manera descrita en el apartado 3.5.2 (porcentaje de identidad mayor del 30%, diferencia en la longitud de las secuencias menor al 10% y diferencia entre las longitudes de las secuencias y el alineamiento menor del 10%).

Una vez identificado los pares de RRSS, para estudiar el tipo de mecanismo evolutivo que se encuentra detrás de cada uno de estos pares es necesario establecer cuando las reacciones del par son metabólicamente cercanas o lejanas. He fijado el umbral para decidir si dos reacciones son cercanas o lejanas en cinco pasos metabólicos basándome en que la mitad de las rutas de biodegradación tienen una longitud menor de cinco pasos siendo además estas rutas las que tienen más del 50% de sus reacciones con un metaenzima asociado. De esta forma cuando la distancia entre el par de reacciones homólogas es menor o igual a cinco pasos, las reacciones se consideran cercanas y si su distancia metabólica es superior a cinco pasos las reacciones se consideran lejanas.

El criterio utilizado para definir el tipo de mecanismo evolutivo asociado a un par de RRSS es el siguiente. Si un par de RRSS están a menos de cinco pasos, se corresponde con un evento de retro-evolución. Si en cambio se encuentran a una distancia superior de cinco pasos, y alguna de sus reacciones vecinas (previas a cada una de las

reacciones del par o posteriores a ellas) constituyen también un par de RRSS, entonces el par se corresponde con un evento de duplicación de rutas. Por último si la distancia entre las reacciones del par es superior a cinco pasos y sus reacciones vecinas no forman otro par de RRSS, es un evento de mosaico.

3.5.6 Tipos de duplicación entre pares de reacciones relacionadas por similitud de secuencia.

Un par de RRSS presentan una duplicación genética si la homología entre las proteínas de sus complejos más informativos cumple que el porcentaje de identidad entre sus secuencias es superior al 90% (para estar muy segura de mantener su actividad catalítica y unión a sustrato (Devos & Valencia, 2000; Devos & Valencia, 2001; Tian & Skolnick, 2003; Hatzimanikatis *et al*, 2004)), además de cumplir que la diferencia de longitudes entre ellas sea menor del 10% y la diferencia entre la longitud del alineamiento y las longitudes de las secuencias no sea superior al 10%. Si el número de proteínas de estos complejos es diferente, es necesario que todas las proteínas del complejo menor tengan una proteína homóloga en el complejo mayor cumpliendo las condiciones anteriores.

Si en cambio, las proteínas de los complejos más informativos de un par de RRSS tienen un porcentaje de identidad de secuencia inferior o igual al 90% (pero siempre superior al 30% y cumpliendo los requisitos de diferencia entre sus longitud de secuencia y con el alineamiento, ya que forman parte de un par de reacciones homólogas), entonces el par de reacciones presentan una dualidad funcional.

3.6 Biología Sintética: El detoxificador de arsénico.

El tipo de modelado utilizado en este trabajo se denomina cinético de ley de masas (*mass-action kinetic*) y consiste en la descripción de los reactantes y productos de cada reacción química como una ecuación donde la velocidad o tasa de la transformación es directamente proporcional a la concentración de los reactantes. De esta manera el circuito del detoxificador de arsénico puede expresarse a través de las siguientes ecuaciones:

Ecuación	Tasa
Transcripción ArsR	
$\text{PlacIq} + \text{RNAP} \rightarrow \text{RNAP} + \text{PlacIq} + \text{mRNA}_{\text{ArsR}}$	k1
Degradación $\text{mRNA}_{\text{ArsR}}$	
$\text{mRNA}_{\text{ArsR}} \rightarrow \text{null}$	k2
Traducción ArsR	
$\text{mRNA}_{\text{ArsR}} + \text{rRNA} \rightarrow \text{mRNA}_{\text{ArsR}} + \text{rRNA} + \text{ArsR}$	k3
Degradación de ArsR	
$\text{ArsR} \rightarrow \text{null}$	k4
Dimerización de ArsR	
$\text{ArsR} + \text{ArsR} \rightarrow \text{ArsR}_2$	k5
$\text{ArsR}_2 \rightarrow \text{ArsR} + \text{ArsR}$	k6
Unión Cooperativa ArsR	
$\text{Pars} + \text{ArsR}_2 \rightarrow \text{ParsArsR}_2$	k8
$\text{ParsArsR}_2 \rightarrow \text{Pars} + \text{ArsR}_2$	k9
Unión de ArsR a As	
$\text{ArsR}_2 + \text{As} \rightarrow \text{ArsR}_2\text{As}$	k11
$\text{ArsR}_2\text{As} \rightarrow \text{ArsR}_2 + \text{As}$	k12
Transcripción l-c-g	
$\text{Pars} + \text{RNAP} \rightarrow \text{RNAP} + \text{Pars} + \text{mRNA}_{\text{l-c-g}}$	k14
Degradación $\text{mRNA}_{\text{l-c-g}}$	
$\text{mRNA}_{\text{l-c-g}} \rightarrow \text{null}$	k15
Traducción de LacI	
$\text{mRNA}_{\text{l-c-g}} + \text{rRNA} \rightarrow \text{mRNA}_{\text{l-c-g}} + \text{rRNA} + \text{LacI}$	k16
Traducción Cyt19	
$\text{mRNA}_{\text{l-c-g}} + \text{rRNA} \rightarrow \text{mRNA}_{\text{l-c-g}} + \text{rRNA} + \text{Cyt19}$	k17
Traducción GFP	
$\text{mRNA}_{\text{l-c-g}} + \text{rRNA} \rightarrow \text{mRNA}_{\text{l-c-g}} + \text{rRNA} + \text{GFP}$	k18
Degradación LacI	

LacI \rightarrow null	k19
Degradación Cyt19	
Cyt19 \rightarrow null	k20
Degradación GFP	
GFP \rightarrow null	k21
Metilación del Arsénico	
Cyt19 + As \rightarrow Cyt19	k22
Tetramerización LacI	
LacI + LacI + LacI + LacI \rightarrow LacI4	k23
LacI4 \rightarrow LacI + LacI + LacI + LacI	k24
Unión cooperativa LacI4	
Plac + LacI4 \rightarrow PlacLacI4	k26
PlacLacI4 \rightarrow Plac + LacI4	k27
Unión cooperativa LacI8	
PlacLacI4 + LacI4 \rightarrow PlacLacI8	k29
PlacLacI8 \rightarrow PlacLacI4 + LacI4	k30
Transcripción de CI	
Plac + RNAP \rightarrow RNAP + Plac + mRNACI	k32
Degradación mRNACI	
mRNACI \rightarrow null	k33
Traducción CI	
mRNACI + rRNA \rightarrow mRNACI + rRNA + CI	k34
Degradación CI	
CI \rightarrow null	k35
Dimerización CI	
CI + CI \rightarrow CI2	k36
CI2 \rightarrow CI + CI	k37
Unión cooperativa CI2	
Pr + CI2 \rightarrow PrCI2	k39
PrCI2 \rightarrow Pr + CI2	k40
Unión cooperativa CI4	
PrCI2 + CI2 \rightarrow PrCI4	k42
PrCI4 \rightarrow PrCI2 + CI2	k43
Transcripción de Pr	
Pr + RNAP \rightarrow RNAP + Pr + mRNAbcx	k45
Degradación mRNAbcx	
mRNAbcx \rightarrow null	k46
Traducción ArsB	
mRNAbcx + rRNA \rightarrow mRNAbcx + rRNA + ArsB	k47
Traducción ArsC	
mRNAbcx + rRNA \rightarrow mRNAbcx + rRNA + ArsC	k48

Traducción XFP	
mRNA _{bcx} + rRNA → mRNA _{bcx} + rRNA + XFP	k49
Degradación ArsB	
ArsB → null	k50
Degradación ArsC	
ArsC → null	k51
Degradación XFP	
XFP → null	k52

Para realizar la simulación del sistema he asignado a cada tipo de reacción un parámetro por defecto cuyo valor he obtenido tras el análisis de los valores utilizados en trabajos previos (Basu *et al*, 2004; Feng *et al*, 2004; Basu *et al*, 2005; Hooshangi *et al*, 2005). Las tasas se muestran en la siguiente tabla:

Descripción	Valor
Tasa de Degradación	0.0050 segundo ⁻¹
Tasa de Transcripción	0.033 moléculas/segundo
Tasa de Traducción	0.033 moléculas/segundo
Tasa de Dimerización directa	0.00001 moléculas/segundo
Tasa de Dimerización inversa	0.20 segundo ⁻¹
Unión cooperativa directa	0.5 moléculas/segundo
Unión cooperativa inversa	0.0010 segundo ⁻¹
ArsR unión a Arsénico directa	0.1 moléculas/segundo
ArsR unión a Arsénico inversa	0.0010 segundo ⁻¹

Los valores iniciales de los elementos del modelo son todos 0 moléculas excepto para los promotores cuyo cantidad inicial es de 1 molécula, la RNAP que tiene 5 moléculas y el rRNA con otras 5 moléculas (aunque al haber cuatro promotores con 4 moléculas es suficiente). Para realizar la simulación he utilizado el paquete de Matlab llamado “SimBiology” (MathWorks, Inc.).

4 RESULTADOS

4.1 Conceptos importantes de la red de Reacciones de Biodegradación.

La red global de biodegradación que se utiliza en este trabajo es una red donde los nodos son las reacciones y las conexiones los compuestos (es decir, una red de reacciones). Dos reacciones se conectan si un producto de la primera reacción es sustrato de la segunda. A consecuencia de esto, las conexiones entre reacciones son dirigidas, proporcionando dirección a la red. Al mismo tiempo, debido a que muchas rutas de biodegradación producen un compuesto final que pertenece al metabolismo central, se ha introducido en la red un nodo artificial que representa la entrada a la red del Metabolismo Central. Utilizando este nodo artificial y la dirección de los enlaces, podemos cuantificar para las reacciones que tienen un camino hacia el Metabolismo Central, a qué distancia se encuentran de éste.

La estructura y relación de los conceptos enzimáticos que describen los procesos de biodegradación es la siguiente (Figura 7). Las proteínas se agrupan en complejos enzimáticos (complejos de proteínas) para poder realizar una reacción. Cada proteína tiene un papel dentro de la reacción, y solo cuando sus acciones se combinan puede llevarse a cabo. En muchas ocasiones una misma reacción puede ser realizada por complejos de proteínas distintos que pertenecen a diferentes organismos o que siendo del mismo organismo no tienen las mismas proteínas. Por ello he llamado metaenzima a la abstracción de esos complejos, es decir, una metaenzima englobaría todos los complejos de proteínas distintos capaces de realizar la misma reacción. Por último, estas reacciones forman parte de rutas de biodegradación.

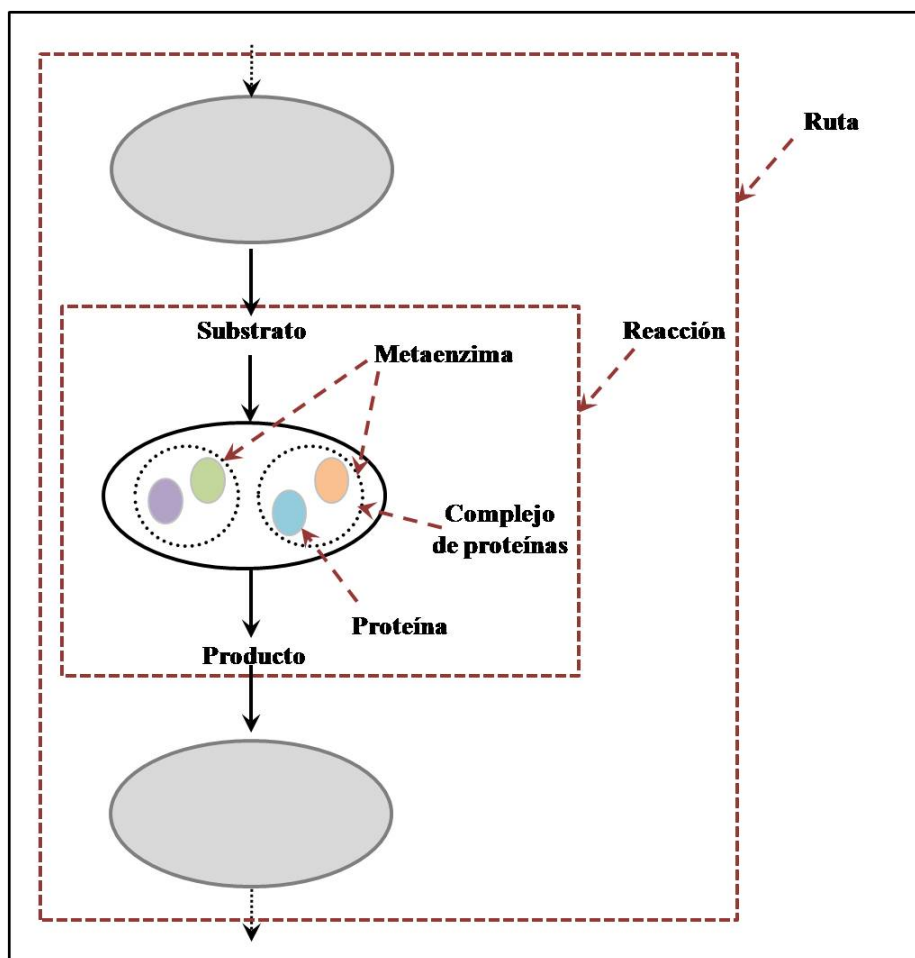


Figura 7. Conceptos enzimáticos y sus relaciones.

Aunque los complejos de proteínas que realizan la misma reacción de biodegradación son homólogos, he encontrado dos casos en que no lo son. La escasez de estos eventos de convergencia funcional (la misma función en secuencias no homólogas) es lógica al tratarse de un evento bastante raro en la evolución de las proteínas (Doolittle, 1994). Uno de los casos de convergencia funcional en la red de biodegradación se da en la reacción que transforma el cloroacetato en glicolato. Esta transformación puede llevarse a cabo por dos dehalogenasas haloacídicas que no son homólogas entre ellas, una de *Alcaligenes xylosoxydans* (*dhlB*) y otra de *Xanthobacter autotrophicus* (*dhlB*) (van der Ploeg *et al*, 1991; Brokamp *et al*, 1996). El otro caso de convergencia se da con dos arsenato reductasas (*arsC*), una en *Escherichia coli* y otra en *Staphylococcus aureus* que sin ser

homólogas entre sí transforman el arsenato en arsenito (Chen *et al*, 1986; Gill *et al*, 2005).

4.2 La base de datos BioNeMo y sus herramientas.

4.2.1 Base de Datos.

El conocimiento de los mecanismos que llevan a la formación y evolución de la red de biodegradación requiere recopilar y relacionar toda la información sobre biodegradación disponible. Además es necesario estructurarlos de tal manera que permita su fácil manejo.

Por ello, he diseñado e implementado una base de datos relacional llamada BioNeMo (Biodegradation Network Modelling), la cual engloba un importante recopilatorio de información metabólica y de regulación sobre biodegradación, y la dispone de manera adecuada para sacar el máximo conocimiento de sus relaciones. Esta base de datos se ha implementado utilizando PostgreSQL (*Berkeley Software Distribution*). Las unidades que componen la base de datos son tablas que contienen datos y relaciones que los conectan.

La base de datos de la Universidad de Minnesota (UMBBD) (Ellis *et al*, 2006) es un recopilatorio muy completo en cuanto a la biodegradación de compuestos recalcitrantes realizados por bacterias y otros organismos. La característica principal de esta base de datos, es que toda su información es bioquímica y se encuentra estructurada en rutas las cuales describen las transformaciones bioquímicas y los metabolitos intermedios producidos por la degradación de un compuesto inicial. Las rutas finalizan cuando todos los productos de una transformación o bien son compuestos que se encuentran en el Metabolismo Central, o bien son metabolitos ya existentes en otras rutas de biodegradación (si dos rutas tienen una parte común, las dos rutas interseccionan en este punto y la serie de transformaciones solo se encuentra en una de las rutas).

Como es usual en las bases de datos metabólicas, las reacciones en UMBBD se describen mediante su número enzimático EC (en inglés “Enzyme Commission number”). El número EC es un esquema de clasificación numérica para las enzimas, basado en las reacciones químicas que catalizan (su actividad bioquímica). Cada código EC consta de 4 números separados por puntos. Estos números representan una clasificación progresivamente más específica, desde el tipo de reacción indicada por el primer dígito (oxidoreductasas, transferasas,

hidrolasas, liasas, isomerasas, y ligasas) hasta el sustrato que transforma representado por el último dígito.

El número EC no es suficiente para identificar la entidad biológica (el complejo proteico) que llevaría a cabo la reacción. Prueba de ello es que más del 60% de las reacciones en biodegradación comparten su código enzimático pero sin embargo la mayoría de ellas no pueden ser realizadas por el mismo complejo proteico ni realizan exactamente la misma transformación. Este hecho se debe principalmente a la ambigüedad a la hora de definir estos códigos y a los diferentes criterios usados por los anotadores cuando fueron asignados.

Para solventar este problema, manteniendo las rutas y las reacciones descritas en UMBBD, he asignado al mayor número posible de reacciones la entidad biológica encargada de realizarla. Esta información se ha obtenido mediante búsquedas manuales en la base de datos de secuencia y en la base de datos bibliográfica del NCBI (Entrez). El resultado es la identificación para cada reacción de las proteínas, los complejos enzimáticos y de lo que he denominado metaenzima.

Las proteínas realizan una actividad dentro de la transformación (actividad catalítica, unión a cofactor, etc.). Cuando ha sido posible, he asignado su código Uniprot (Swissprot y TrEMBL), su código NCBI, sus dominios funcionales descritos en la base de datos Pfam, su estructura tridimensional en la base de datos PDB, el gen (con su código NCBI) que codifica para la proteína y su organismo (con su código NCBI).

Teniendo en cuenta los conceptos de proteína, complejo de proteínas, enzima y reacción definidos en la descripción de la red de reacciones, la estructura de la información enzimática en BioNeMo se dispone de la siguiente manera (Figura 8). Las rutas de biodegradación contienen reacciones, cada reacciones tienen complejos de proteínas capaces de realizarlas, cada complejo de proteínas tiene asociado unas proteínas y cada proteína cuenta con una gran cantidad de información tanto en la propia base de datos de BioNeMo como gracias al gran número de enlaces a bases de datos públicas y de interés.

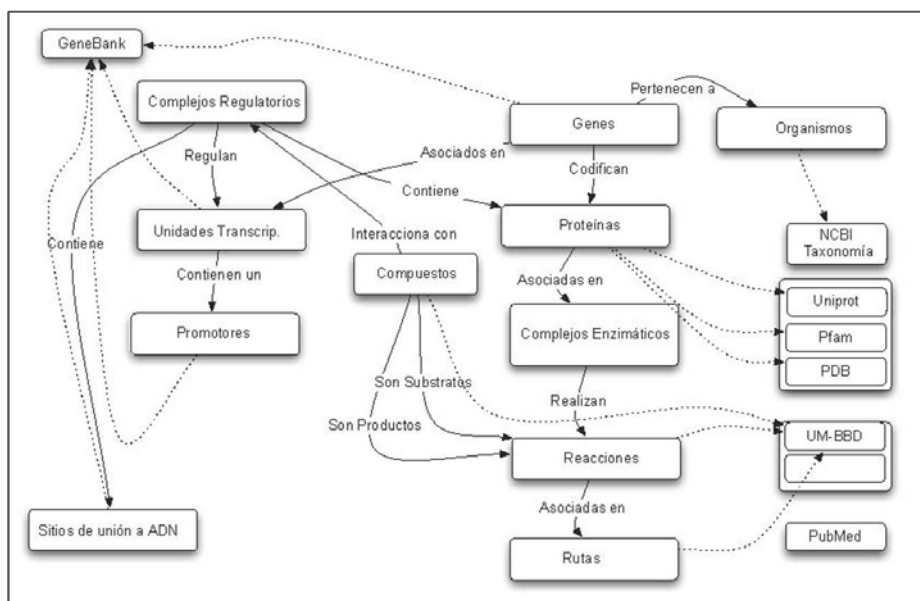


Figura 8. Relaciones en BioNeMo. La figura muestra las conexiones entre todas los elementos tanto enzimáticos como de regulación contenidos en BioNeMo y sus conexiones con bases de datos externas. Las flechas continuas representan relaciones internas en BioNeMo y las flechas discontinuas relaciones con bases de datos externas.

La metodología utilizada en la obtención de estos datos se explica en la sección de Materiales y Métodos, sin embargo creo importante resaltar la gran inversión de tiempo dedicado en su realización. Como resultado de este esfuerzo, la colección de datos enzimáticos se muestra en la siguiente tabla:

Entidad	Número
Rutas	146
Reacciones	945
Complejos enzimáticos	533
Proteínas	836
Secuencia de proteínas	779
Dominios funcionales	192
Estructuras tridimensionales	332
Microorganismos	195

Gracias a toda esta información, la base de datos BioNeMo es una herramienta muy útil para obtener nuevo conocimiento sobre biodegradación a partir de la relación y el análisis de los datos

asociados a un gran número de reacciones de biodegradación. Sin embargo dada la complejidad de su estructura, su manejo no resulta fácil salvo que se tenga experiencia realizando consultas mediante el lenguaje de bajo nivel SQL (*Structured Query Language*) y se conozca perfectamente las relaciones entre sus tablas.

4.2.2 Interfaz de programación de aplicaciones (API).

Para solventar dicho problema y evitar que las consultas a la base de datos sean tediosas, se ha desarrollado e implementado una interfaz de programación de aplicaciones (API) utilizando el lenguaje de programación Perl y una interfaz *web*.

Un API (del término inglés *Application Program Interface*) es el conjunto de funciones y métodos que ofrece una cierta librería para ser utilizado por otro software como una capa de abstracción. Uno de los principales propósitos de una API consiste en proporcionar un conjunto de funciones de uso general, de forma que los usuarios (programadores) pueden hacer uso de su funcionalidad evitándose programar todo desde el principio.

La API que hemos desarrollado engloba un conjunto de paquetes los cuales se corresponden con cada una de las entidades que se encuentran en BioNeMo: rutas, reacciones, complejos, proteínas, secuencias de proteínas, genes, dominios funcionales (Pfam), estructuras 3D (Pdb) y organismos. Cada uno de estos paquetes contienen métodos que son consultas más o menos sencillas a la base de datos acerca de las propiedades de esas entidades. Enlazando esos métodos es posible realizar consultas muy complejas utilizando un código de programación muy lógico y estructurado.

Un ejemplo que lo ilustra es el siguiente. Si quisiera saber las secuencias de las proteínas que participan en las reacciones de la ruta del “Acrilonitrile”, la consulta SQL a la base de datos sería la siguiente:

```
SELECT sp.sequence FROM sequence_protein sp, protein_sequence_protein
psp, protein p, complex_protein cp, reaction_complex rc,
pathway_reaction pr, pathway pa WHERE pa.name = 'Acrylonitrile' AND
pa.id_pathway = pr.id_pathway AND pr.id_reaction = rc.id_reaction AND
rc.id_complex = cp.id_complex AND cp.id_protein = p.id_protein AND
p.id_protein = psp.id_protein AND psp.id_sequence = sp.id_sequence;
```

Sin embargo utilizando la API, la consulta es:

```
$pathway_object = $conexobj->get_path( description,'Acrylonitrile' );
$pathway_object->[0]->display_description('text');
$array_reaction_objects = $pathway_object->[0]->its_reactions;

foreach my $reaction_object (@$array_reaction_objects) {
    $array_complex_objects = $reaction_object->its_complexes;
    foreach my $complex_object (@$array_complex_objects) {
        $array_protein_objects = $complex_object->its_proteins;
        foreach my $protein_object (@$array_protein_objects) {
            print ">";
            $protein_object->its_sequence->display_sequence('flat');
        }
    }
}
```

El acceso a la base de datos de este modo resulta mucho más intuitivo, no precisa de un conocimiento exhaustivo de las tablas implicadas en la consulta ni de los campos dentro de las tablas. Simplemente sabiendo que una ruta tiene reacciones, las reacciones tienen complejos enzimáticos, los complejos enzimáticos están formados por proteínas, y las proteínas tienen secuencia, se consigue realizar la consulta.

4.2.3 Servidor Web.

La segunda herramienta desarrollada es una interfaz web de la base de datos BioNeMo. Dado que BioNeMo puede ser de gran utilidad para laboratorios trabajando en biodegradación, esta aplicación permite consultarla sin necesidad de programar.

El servidor muestra la información y sus relaciones en la base de datos utilizando diferentes tipos de páginas y enlazando/cruzando su información. Existe una página inicio, una de resultados “google” (que se explicará luego), una para mostrar la información acerca de la ruta, una para mostrar la información acerca de la reacción y su complejo-enzima (el conjunto de complejos de proteínas que la realizan), una para los datos del gen y por último otra con la información de la unidad transcripcional. La relación entre todas ellas se muestra en la Figura 9. Explicaré las más relevantes y relacionadas con la actividad metabólica, no la regulación, ya que no es el tema de esta tesis y ha sido realizado por Guillermo Carbajosa.

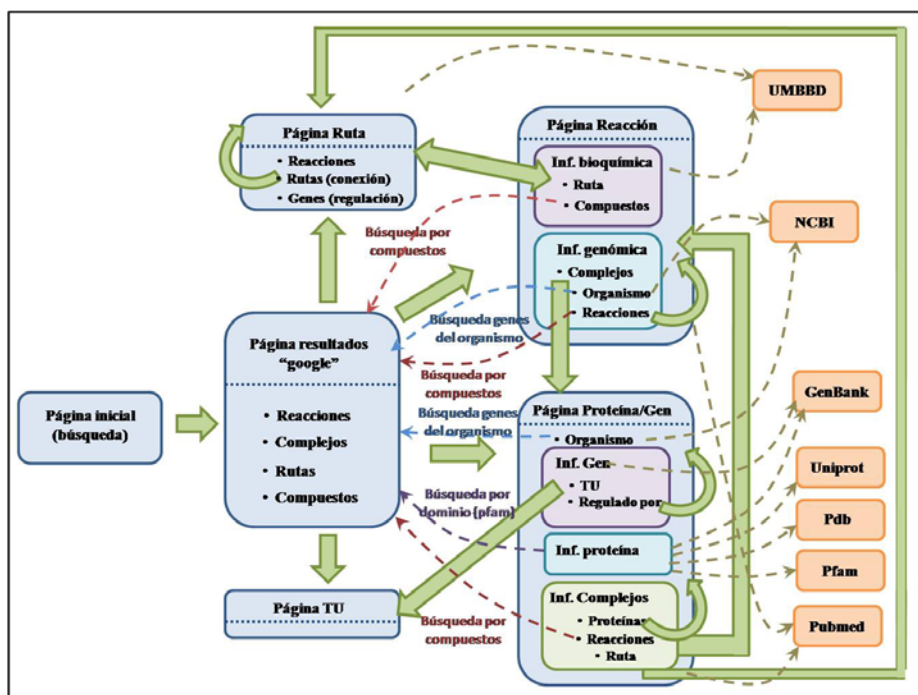


Figura 9. Estructura del servidor web de BioNeMo. Las flechas verdes gruesas indican transiciones entre páginas, las flechas verdes discontinuas enlaces con bases de datos externas, las flechas discontinuas marrones búsquedas "tipo google" a partir del nombre de un compuesto y las flechas discontinuas azules búsquedas de los genes que tiene un organismo. Los bloques azules hacen referencia a las páginas del servidor web y los naranjas a las bases de datos externas. Los bloques (morados, azul claro y amarillos) dentro de las páginas indican el diferente tipo de información que se puede mostrar en ellas.

La página inicio permite introducir un texto a buscar (Figura 10). Su funcionamiento se asemeja al motor de búsqueda "google", es decir, muestra como resultado todo lo que dentro de la base de datos contenga o se corresponda con el texto introducido.

Bionemo

Search:

Biodegradation Network Modelling Database

[Contact](#)

Bionemo is a database of Genes related to biodegradation and their Regulation. It complements other biodegradation Databases as [MetaRouter](#) or [The University of Minnesota Biocatalysis/Biodegradation Database](#). While these databases have focused in the biochemistry of the biodegradation processes, Bionemo is focused in Genes and Proteins.

We have collected in **Bionemo** all the DNA and protein sequences available for biodegradation enzymes and enzyme units. The organization of enzymatic complexes and transcription units is also stored. Transcription Factors and their target genes and binding sites where manually retrieved from the Literature and included in **Bionemo**.

To start using Bionemo, just type your query here:

examples: xylR, phenol, dioxygenase

You can also download the whole relational database. To help you accessing the data, we also provide a Perl API.

What's in Bionemo:

Metabolic Genes Their DB Entries (sequence, domains, structure), Enzymatic complexes they belong, how they are transcribed and their regulation.

Transcription Factors Which genes they regulate and which effectors they interact

Transcription Units Which genes they contain, which transcription factors regulate them, and where they bind

Metabolic Reactions Which complexes perform them, and which genes code for them, in which species they appear

You can use **Pathways** and **Compounds** to navigate this information.

Downloads:

[PostgreSQL Dump](#)

[Bionemo Perl API](#)

[All Protein Sequences in a Single FASTA file](#)

Figura 10. Página de inicio del servidor web BioNeMo.

Los resultados de la búsqueda se muestran clasificados en genes, unidades transcripcionales, reacciones, complejos de proteínas, rutas, organismos y compuestos (Figura 11). Cada uno de estos resultados enlaza con una serie de páginas o sucesivas búsquedas “tipo google” para refinar el resultado. A continuación explicaré aquellas que son más relevantes a nivel metabólico.

Bionemo Search:

Biodegradation Network Modelling Database [Contact](#) [Home](#)

We found:

Reactions(16)	Complexes(14)	Pathways(22)	Compounds(24)
---------------	---------------	--------------	---------------

Complex: [benA](#) [benB](#) [benC](#) [Acinetobacter sp. ADP1](#)
 Reaction: [o-Methylbenzoate](#) --> [1,2-Dihydroxy-6-methylcyclohexa-3,5-dienecarboxylate](#)
 Reaction: [m-Methylbenzoate](#) --> [1,2-Dihydroxy-3-methylcyclohexa-3,5-dienecarboxylate](#)
 Reaction: [Benzoate](#) --> [cis-1,6-Dihydroxy-2,4-cyclohexadiene-1-carboxylicacid](#)

Complex: [cbdA](#) [cbdB](#) [cbdC](#) [Burkholderia sp. TH2](#)
 Reaction: [2-Chlorobenzoate](#) --> [CarbonDioxide](#) + [Catechol](#)

Complex: [badA](#) [Rhodopseudomonas palustris CGA009](#)
 Reaction: [Benzoate](#) --> [Benzoyl-CoA](#)

Complex: [ohbA](#) [ohbB](#) [Pseudomonas aeruginosa PAC433](#)
 Reaction: [2-Chlorobenzoate](#) --> [CarbonDioxide](#) + [Catechol](#)

Complex: [cbdB](#) [cbdC](#) [cbdC](#) [Burkholderia cepacia](#)
 Reaction: [2-Chlorobenzoate](#) --> [CarbonDioxide](#) + [Catechol](#)

Complex: [pobB](#) [Pseudomonas sp.](#)
 Reaction: [4-Hydroxybenzoate](#) --> [3,4-Dihydroxybenzoate](#)
 Reaction: [4-Hydroxybenzoate](#) --> [3,4-Dihydroxybenzoate](#)

Complex: [bclA](#) [Azoarcus sp. EbN1](#)
 Reaction: [Benzoate](#) --> [Benzoyl-CoA](#)

Complex: [hapE](#) [Pseudomonas fluorescens LP6a](#)
 Reaction: [4-Hydroxyacetophenone](#) --> [4-Hydroxyphenyl acetate](#)

Figura 11. Página de resultados "google" (Búsqueda por *Benzoate*). Desde la información proporcionada por cada complejo se puede acceder a la página de sus genes/proteínas y a la de las reacciones que participa. También se puede mostrar todos los genes del organismo al que pertenece.

Una ruta será resultado de la búsqueda si el texto introducido se corresponde con el sustrato o el producto de alguna de sus reacciones. Cada ruta resultado enlaza con las páginas de cada una de esas reacciones y con una vista general de la ruta.

Del mismo modo, una reacción será resultado de la búsqueda "google", si el sustrato o el producto contienen el texto introducido. Desde una reacción resultado se puede navegar hacia la página de la propia reacción y a la página de los genes de los complejos enzimáticos capaces de llevar a cabo esa reacción. También desde los organismos donde se ha descrito que se da dicha reacción se puede realizar una nueva búsqueda y obtener todos los genes que en BioNeMo hay de ese organismo.

Los complejos de proteínas resultado se obtienen si el texto introducido se encuentra presente en el nombre de los genes del complejo o en el nombre de la actividad enzimática que se le ha dado al complejo-enzima que realiza la reacción. Cada complejo encontrado enlaza a la

página de cada una de sus proteínas/genes, a las páginas de las reacciones que realiza ese complejo, y de nuevo se puede realizar una búsqueda a partir del organismo que tiene ese complejo.

Los genes resultado de la búsqueda se obtienen si el texto introducido coincide o está contenido en el nombre del gen. Desde él se puede acceder a la unidad transcripcional (u operón) en el que se encuentra y realizar una búsqueda a partir de su organismo.

Los compuestos resultado se dan porque su nombre contiene o se corresponde con el texto introducido. A partir de él se puede acceder a las páginas de las reacciones en las cuales es sustrato y a las páginas de las reacciones en las cuales es producto. También se puede realizar una nueva búsqueda “tipo google” con el nombre de cualquiera de los compuestos resultado para refinar la consulta.

La página de ruta (Figura 12) ofrece una visión general de la misma y de manera muy visual permite el acceso fácil y directo a toda la información disponible sobre ella. Esta vista diferencia las reacciones a las que he asignado una metaenzima y las reacciones que sólo contienen información bioquímica.

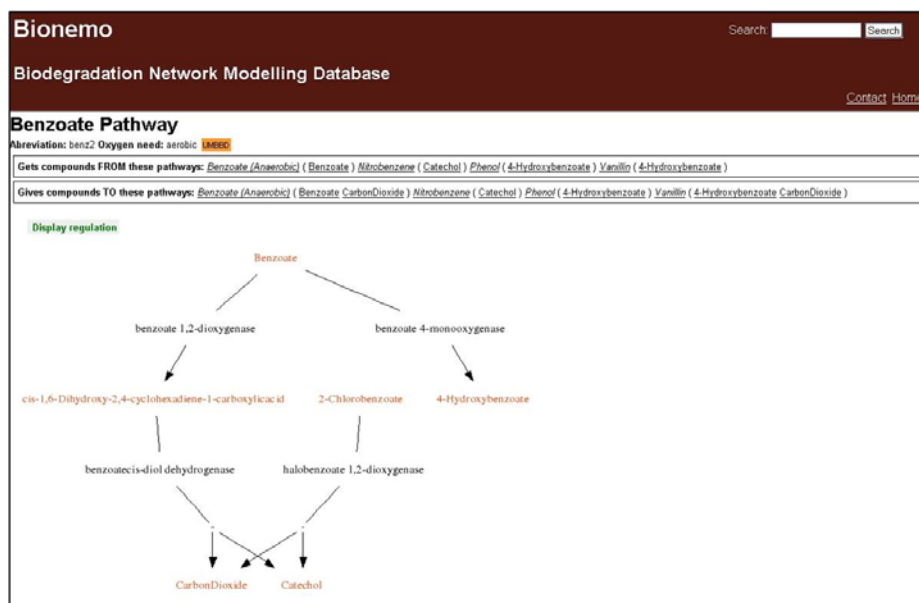


Figura 12. Página de ruta de Benzoato. En naranja se muestran los compuestos, los cuales llevan a una búsqueda tipo "google" a partir de ellos. En negro se representan las reacciones utilizando el nombre que la UMBBD ha asignado a su actividad enzimática.

En las páginas que voy a describir a continuación no se muestra información bioquímica ya que el objetivo de esta base de datos es cubrir la descripción “enzimática” de los procesos de biodegradación, y de esta forma ser complementaria a la información bioquímica existente en UMBBD. Por ello, las páginas de BioNeMo cuentan con un enlace a la base de datos UMBBD.

En la página de reacción (Figura 13) se muestra tanto la información bioquímica (substratos, productos, código enzimático) como la biológica (metaenzima que la realiza). Desde esta página se puede enlazar con la visión general de la ruta a la que pertenece la reacción y con la página de genes que forman parte de cada uno de los complejos enzimáticos de la reacción. También se enlaza con la página de resultados al realizar una búsqueda “tipo google” a partir de los compuestos o al obtener todos los genes del organismo. También enlaza con el artículo donde se describe la transformación bioquímica cuando se dispone de él.

Biodegradation Network Modelling Database

Contact Home

Reaction: 2-Chlorobenzoate ----> CarbonDioxide + Catechol

EC-Code: 1.14.12.13

View more information at: [UMBB](#)

2-Chlorobenzoate ----> CarbonDioxide + Catechol

in: Benzoate Pathway (benz2)

Enzymatic complexes

cbdC cbdB cbdC

Subunits: [CbdB \(sp:CBDB_BURCE\)](#) [CbdC \(sp:CBDA_BURCE\)](#) [CbdC \(sp:CBDC_BURCE\)](#)

From: [Burkholderia cepacia](#) [NCBI](#)

2-Chlorobenzoate ----> CarbonDioxide + Catechol

in: Benzoate Pathway (benz2)

cbdA cbdB cbdC

Subunits: [CbdA \(tr:Q9AJR2_BURST\)](#) [CbdB \(tr:Q9AJR1_BURST\)](#) [CbdC \(tr:Q9AJR0_BURST\)](#)

From: [Burkholderia sp. TH2](#) [NCBI](#)

2-Chlorobenzoate ----> CarbonDioxide + Catechol

in: Benzoate Pathway (benz2)

ohbA ohbB

Subunits: [OhbA \(tr:Q9Z5W2_PSEAE\)](#) [OhbB \(tr:Q9Z5W0_PSEAE\)](#)

From: [Pseudomonas aeruginosa PAC433](#) [NCBI](#)

2-Chlorobenzoate ----> CarbonDioxide + Catechol

in: Benzoate Pathway (benz2)

Articles

[Degradation of 2-chlorobenzoate by Pseudomonas cepacia 2CBS. \(PMD:2610934\)](#)

Figura 13. Página de reacciones. En la parte superior se muestra un pequeños resumen de su información bioquímica, enlazando con la base de datos UMBBD para encontrar información más detallada. En la parte inferior se muestra la información propia de esta base de datos, los complejos asociados a la reacción.

La página de gen/proteína (Figura 14) muestra tanto la información propia del gen y de la proteína que codifica: sus códigos NCBI, Uniprot, sus dominios funcionales en Pfam y su estructura 3D de PDB (siendo todas ellas enlaces a bases de datos externas), como la información acerca de su complejo enzimático al que pertenece y reacciones que realiza. Al igual que antes se puede enlazar con la página de resultados al obtener todos los genes en BioNeMo que comparten alguno de sus dominios funcionales, al obtener todos los genes del organismo que tiene este gen, o al realizar una búsqueda “tipo google” a partir de los compuestos de las reacciones en las que participa.

Bionemo Search: Search

Biodegradation Network Modelling Database [Contact](#) [Home](#)

Gene: benC

Source: [Acinetobacter sp. ADP1](#) [NCBI](#)

DNA: [CR543861.1](#)

Transcriptional units: [benABCDE](#) [operator5]

Regulated by [benM](#) [catM](#)

Protein: BenC ([sp|BENC_ACIA0](#), [CAG68302](#))

Pfams:

[2Fe-2S iron-sulfur cluster binding domain \(PF00111\)](#) [Pfam] Start: 19 End: 99

[Oxidoreductase FAD-binding domain \(PF00970\)](#) [Pfam] Start: 120 End: 215

[Oxidoreductase NAD-binding domain \(PF00175\)](#) [Pfam] Start: 224 End: 328

Pdbs

[1KRH_A](#) [1KRH_B](#)

Enzymatic complexes

benA benB benC

m-Methylbenzoate → **1,2-Dihydroxy-3-methylcyclohexa-3,5-dienecarboxylate**
in: [m-Xylene Pathway \(mxy\)](#)

o-Methylbenzoate → **1,2-Dihydroxy-6-methylcyclohexa-3,5-dienecarboxylate**
in: [o-Xylene Pathway \(oxy\)](#)

Benzoate → **cis-1,6-Dihydroxy-2,4-cyclohexadiene-1-carboxylic acid**
in: [Benzoate Pathway \(benz2\)](#)

No **Pubmed** found

[Top](#)

Figura 14. Página gen/proteína. En la parte superior se comienza mostrando información acerca del gen, para seguir con la información relacionada con la proteína que codifica, y terminar con el complejo de proteínas y las reacciones en las que participa.

Los datos contenidos en BioNeMo suponen un complemento respecto a la información que se tenía de biodegradación debido por una parte a la asignación de la entidad biológica a un gran número de reacciones o

transformaciones bioquímicas, a la conexión entre las actividades metabólicas y de regulación y a la organización de estos datos en una estructura que permita el estudio de grandes cantidades de información y de sus relaciones. A su vez, el desarrollo de la API facilita enormemente el manejo de estos datos y permite su utilización para realizar estudios computacionales sin tener un conocimiento exhaustivo de la estructura de la base de datos. Por último, el diseño y creación de la interfaz web amplia su utilización, ya que permite realizar consultas a la base de datos de forma muy sencilla, visual y sin necesidad de programar. En la programación de la interfaz web se ha utilizado la API previamente desarrollada.

4.3 Grado de descripción de la red de Biodegradación.

Una vez que la colección de datos en BioNeMo estaba lista para ser analizada, necesité comprobar si el nivel de descripción de la red de biodegradación era suficiente para poder sacar conclusiones sobre ella. El nivel de descripción de la red se definirá en función del grado de información que tenga cada una de sus reacciones y del porcentaje de reacciones en la red que tengan ese grado de información. Se puede considerar que la descripción molecular de las reacciones se divide en cuatro niveles (Figura 15).

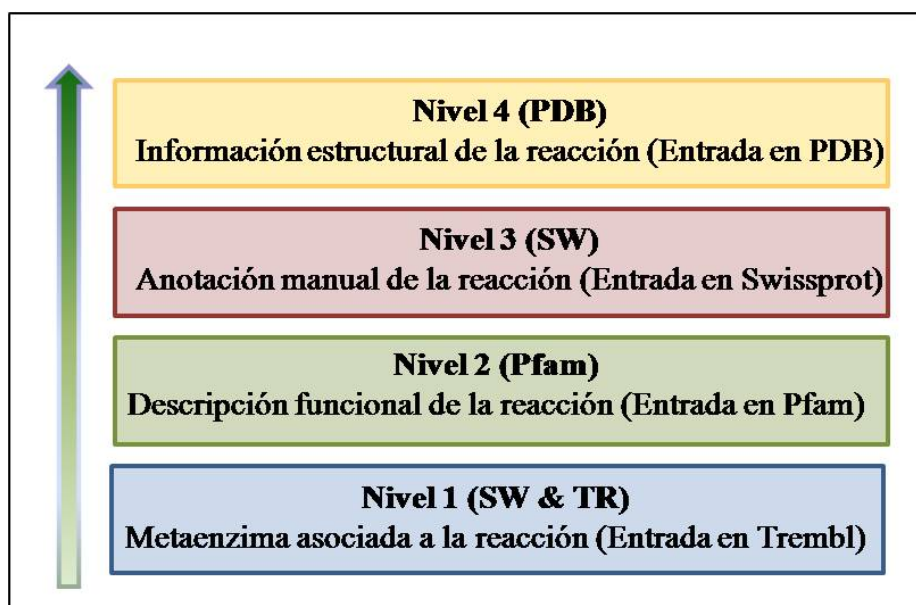


Figura 15. Definición de los niveles de caracterización de las reacciones. En la figura el nivel azul representa el nivel inferior de descripción y el amarillo el mayor.

El nivel inferior de descripción (nivel 1) se corresponde con la existencia de una metaenzima asociada a la reacción. Como ya ha sido mencionado, se ha obtenido tras una asignación manual al mayor número de reacciones posible de la metaenzima que realiza la transformación. Como todas las secuencias de la metaenzima tienen una entrada en la base de datos de secuencias Trembl, me refiero a este nivel con las siglas “SW & TR”. Trembl (del inglés “Translation of EMBL”) es una base de datos biológica de secuencias de proteínas que

incluye la traducción computacional de todas las secuencias codificantes derivadas del (EMBL-BANK) y que todavía no han podido ser anotadas manualmente. Su anotación se realiza de forma automática e incluye proteínas marcadas como hipotéticas, fragmentos, etc. (Boeckmann *et al*, 2003).

Sólo para las reacciones que tienen secuencia, se puede ver si alcanzan un nivel de descripción mayor (nivel 2) en el cual se estudia la descripción funcional de la reacción. Esta información se obtiene buscando dominios funcionales en la metaenzima que realiza las reacciones usando la base de datos de dominios Pfam. Esta base de datos contiene una amplia colección de alineamientos múltiples y modelos ocultos de Markov que cubren los dominios de las proteínas y familias más comunes cuya información funcional es disponible (Finn *et al*, 2007). De este modo, si la metaenzima de las reacciones de biodegradación contiene alguno de los dominios descritos en Pfam, se puede decir que la reacción tiene función aproximada conocida, sin embargo si no se encuentra información en Pfam, la descripción funcional de la reacción no se asemeja a ninguna función previamente descrita.

El tercer nivel de descripción se corresponde con la anotación específica de la metaenzima de las reacciones. Mientras que tanto en el nivel 1 como en el nivel 2, los métodos utilizados eran computacionales, este nivel se basa en la anotación manual de la metaenzima. La revisión manual de la información aporta una mayor confianza a los datos. SwissProt es una base de datos biológica de secuencias de proteínas. Fue creada en 1986 por Amos Bairoch durante su tesis doctoral y desarrollada por el instituto Suizo de Bioinformática y el Instituto Europeo de Bioinformática. Su característica principal es que las proteínas que se encuentran en esta base de datos tienen una anotación muy detallada, es decir, cada secuencia ha sido revisada, documentada y enlazada a otras bases de datos manualmente (Boeckmann *et al*, 2003). Únicamente las proteínas que son revisadas manualmente son ascendidas de Trembl a SwissProt, por tanto la presencia de la metaenzima en esta última es un indicador de la fiabilidad de la información que se dispone de la reacción.

El nivel más alto de descripción (nivel 4) se corresponde con la información estructural de la reacción. En general se considera que la disponibilidad de la estructura tridimensional (3D) es la caracterización molecular más detallada posible ya que la determinación de la misma

suele ser el último paso tras una completa caracterización bioquímica y funcional. La estructura 3D de la metaenzima ha sido obtenida del repositorio de estructuras de proteínas PDB (en inglés “Protein Data Bank”). Esta base de datos fue creada en 1971 por Edgar Meyer and Walter Hamilton y más tarde transferida a los miembros del RCSB (“Research Collaboratory for Structural Bioinformatics”). Los datos normalmente se obtienen de cristalografía rayos X o espectroscopia NMR (Wüthrich, 1986).

4.3.1 La red de Biodegradación Vs la red del Metabolismo Central.

La red del Metabolismo Central (CM) ha sido estudiada de forma exhaustiva, siendo la descripción que de ella se tiene muy completa. Además toda esta información ha sido dispuesta en una estructura computacional desde la que se pueden realizar estudios sistemáticos de la red. Por ello utilizaré la red del Metabolismo Central como base para comparar el nivel de caracterización disponible para la red de biodegradación.

A todos los niveles de descripción, la red del Metabolismo Central está mejor descrita que la red global de biodegradación. Mientras que en los niveles más bajos (la asociación de un complejo-enzima a la reacción y su descripción funcional) la descripción de la red del Metabolismo Central es un 75% superior en ambos casos, según se incrementa el nivel de información (se pasa de métodos computacionales a manuales), esta diferencia es aún mayor, siendo la anotación manual y la caracterización estructural de las reacciones de la red del Metabolismo Central dos veces y cuatro veces, respectivamente, superior (Figura 16).

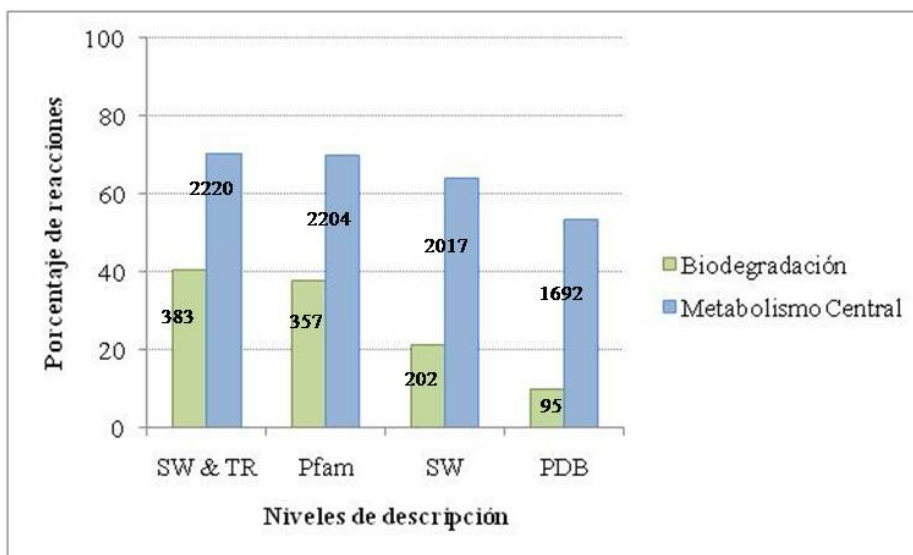


Figura 16. Comparación del grado de descripción de las reacciones en la red de Biodegradación y en la del Metabolismo Central clasificada en niveles. El nivel más bajo se corresponde con la asignación de secuencia a la reacción ("SW & TR"), seguido de la asignación funcional de manera automática "Pfam", de su anotación manual "SW", y por último la asignación de una estructura tridimensional "PDB". El número de reacciones para cada nivel aparece sobre la figura, siendo el número total de reacciones en biodegradación 945 y en el Metabolismo Central 3155. Los valores sobre la figura son número de casos, no porcentajes.

4.3.2 Biodegradación total Vs. biodegradación parcial.

Dentro de la red de biodegradación se pueden diferenciar las reacciones que participan en la completa biodegradación de los compuestos y las que no. En términos de red, que una reacción participe en la degradación total de un compuesto quiere decir que conecta con el Metabolismo Central.

Las reacciones conectadas con el Metabolismo Central se conocen mejor a todos los niveles. A nivel de secuencia un 50% más de reacciones que llegan al CM tienen un complejo-enzima asociado respecto de las que no llegan. De igual modo, un 30% más de reacciones que conectan al CM tienen descripción funcional, alrededor de un 45% más de reacciones que conectan con el CM han sido anotadas manualmente y un 42% más de reacciones que conectan con el CM tienen estructura tridimensional. En este caso, la diferencia entre

la descripción de los dos tipos de reacciones no incrementa conforme aumenta el nivel de información (Figura 17).

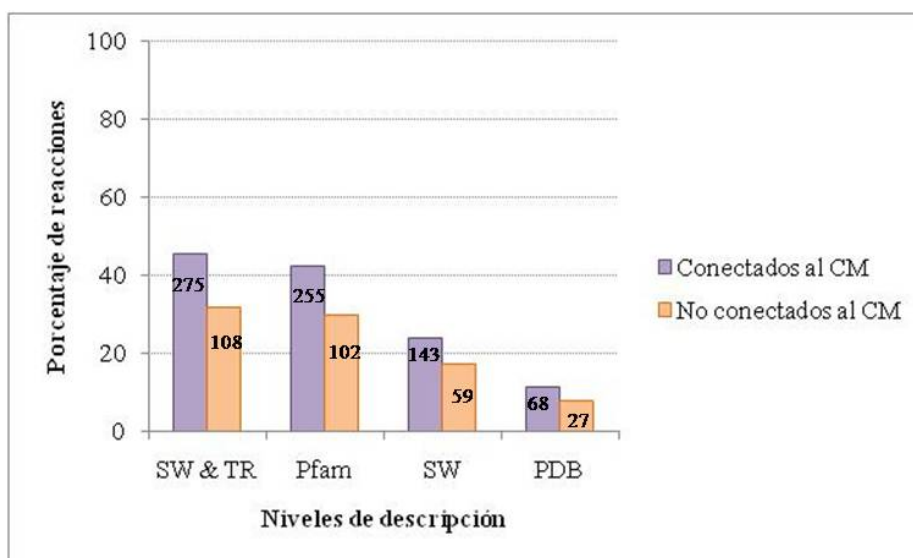


Figura 17. Comparación del grado de descripción de las reacciones que forman parte de procesos de biodegradación total respecto a las reacciones que participan en biodegradaciones parciales. Los números sobre las barras indican el número de reacciones en cada nivel. El número total de reacciones conectadas al CM es 604 y no conectadas al CM 341. Los valores sobre la figura son número de casos, no porcentajes.

4.3.3 Caracterización a lo largo de la red de Biodegradación.

Incorporando el concepto de distancia al CM en el análisis, se puede observar la distribución de la descripción a lo largo de la red de biodegradación (periférica, zona intermedia y próxima al CM). De esta manera es posible identificar si existen regiones de la red más estudiadas, tendencias respecto a la situación o la manera en la que la red ha sido experimentalmente analizada.

Tanto para la asociación de complejo-enzima a las reacciones (45%, 50% y 45% a las distancias 1 y 2, 3 y 4 y mayor 4, respectivamente) como para la descripción funcional de las reacciones (42%, 45% y 40% a las distancias 1 y 2, 3 y 4 y mayor 4, respectivamente) y para su anotación manual (25%, 25% y 25% a las distancias 1 y 2, 3 y 4 y

mayor 4, respectivamente), el porcentaje de reacciones se mantiene prácticamente constante a lo largo de toda la red (Figura 18).

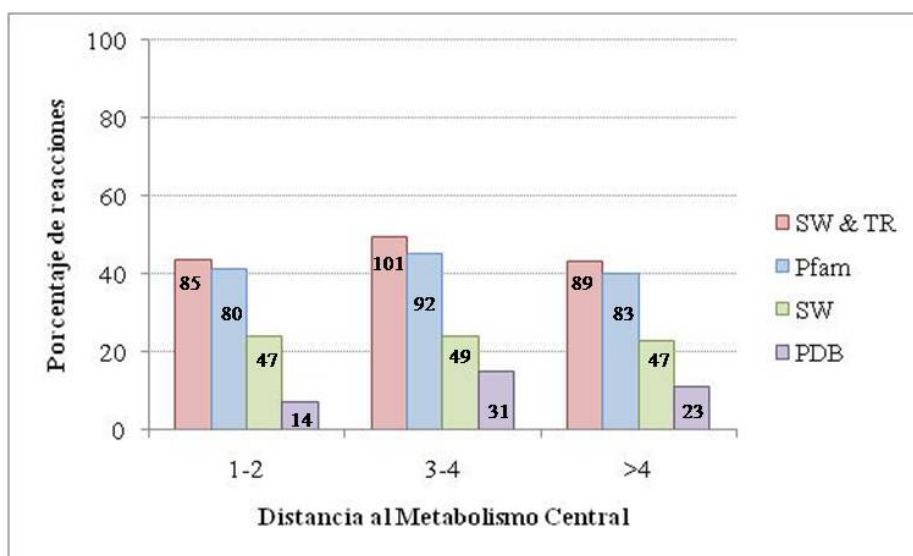


Figura 18. Grado de descripción de las reacciones de la red de biodegradación que participan en procesos de biodegradación total, clasificados según el nivel de caracterización y la distancia de las reacciones a la entrada a la red del Metabolismo Central. El número total de reacciones a distancia 1 y 2 es 194, a distancia 3 y 4 es 204 y a distancias mayores de 4 es 206. Los valores sobre la figura son número de casos, no porcentajes.

Los resultados obtenidos indican que la red de biodegradación ha sido mucho menos estudiada que la del Metabolismo Central y que las reacciones que participan en la biodegradación completa de los compuestos se han caracterizado mejor que las que no consiguen degradarlos totalmente hasta su entrada al CM. Sin embargo, la información que se tiene de la red de biodegradación está homogéneamente distribuida y por tanto aunque es necesario incrementar los esfuerzos para mejorar su descripción, podemos utilizarla para identificar patrones o tendencias en su organización, formación, y evolución.

En la Figura 19 se muestra una representación de la red Global de Biodegradación, con el grado de caracterización de cada una de sus reacciones.

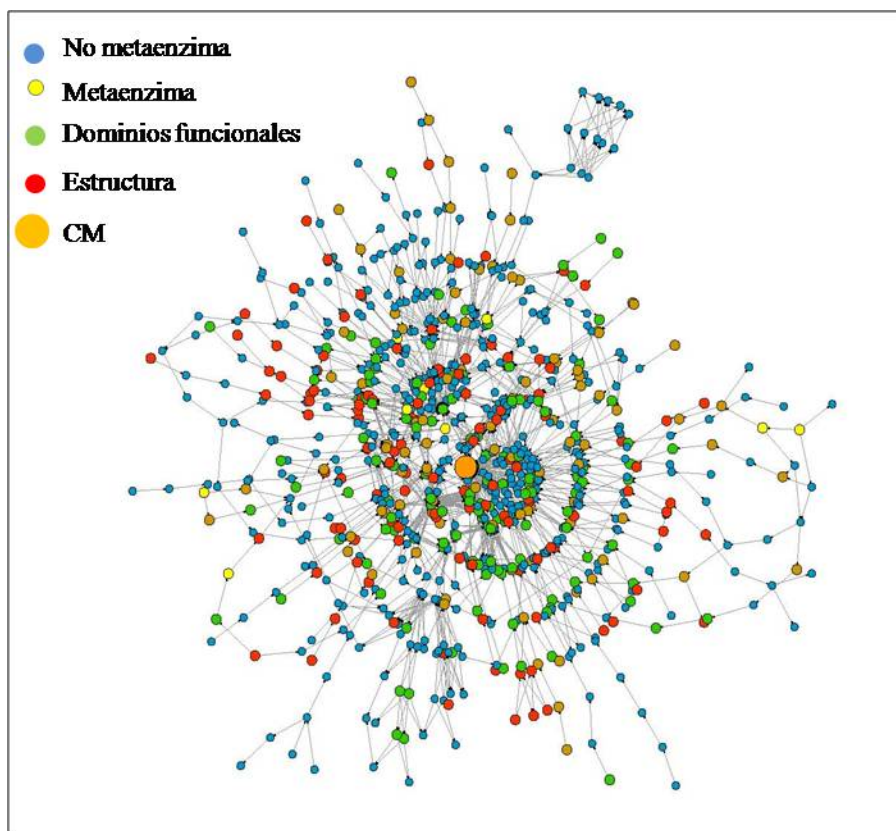


Figura 19. Visión global de la red de reacciones de biodegradación. Los nodos en azul se corresponden con reacciones a las que no ha sido posible la asignación de la metaenzima, los amarillos las reacciones a las que como máximo he podido asignar la metaenzima, los verdes las reacciones que también tienen descripción de sus dominios funcionales y en rojo las reacciones con incluso estructura 3D. El nodo central naranja representa la entrada al Metabolismo Central.

Prácticamente la mitad de las reacciones con una metaenzima asociada se encuentran o completamente aisladas, es decir, las reacciones vecinas no tienen metaenzima (42 casos) o formando grupos de dos o tres elementos (140 casos). Las 201 reacciones restantes pertenecen a una subred interconectada.

4.4 Estudio topológico de la red de Biodegradación.

Como ya he mencionado, lo que diferencia a este trabajo de lo realizado anteriormente es la combinación de un análisis global de la biodegradación, a diferencia de los numerosos estudios sobre rutas específicas, la consideración de la biodegradación como un evento global independiente del organismo, a diferencia de los numerosos estudios de rutas en organismos específicos, y su enfoque molecular más que bioquímico con la asignación de una metaenzima a las reacciones en lugar del uso del número enzimático EC y utilización de una red de reacciones en lugar de una red de compuestos. La red que se obtiene de esta combinación, aun siendo metabólica, puede presentar una topología diferente de las observadas en la red de compuestos del Metabolismo Central (Jeong *et al*, 2000; Wagner & Fell, 2001; Ravasz *et al*, 2002) y en la red de compuestos de biodegradación (Pazos *et al*, 2003).

4.4.1 Estructura libre de escala.

La primera de las características topológicas que han sido estudiadas en ambas redes es la distribución del número de conexiones de sus nodos. Como las redes metabólicas son dirigidas, un nodo presenta dos tipos de conectividades: conexiones de entrada y conexiones de salida.

Se ha visto que las redes metabólicas presentan una estructura libre de escala, es decir, mientras la mayoría de los nodos tienen muy pocas conexiones, un número pequeño tienen una conectividad muy alta. En este tipo de redes la probabilidad $p(k)$ de que un nodo de la red esté conectado con k nodos es proporcional a $k^{-\gamma}$, es decir, sigue una ley de potencias. El exponente γ no es universal, sino que depende del tipo específico de la red y para los sistemas biológicos debería encontrarse en el rango $2 < \gamma < 3$.

El estudio realizado por Jeong *et al*. para la red metabólica de 43 organismos, revela que la red de compuestos del Metabolismo Central para todos ellos tenía una estructura libre de escala con un exponente alrededor de 2,2 (Jeong *et al*, 2000). Por su parte la red de compuesto de biodegradación analizada por Pazos *et al*. también presenta una estructura libre de escala con un exponente aproximadamente entre 2 y

3 tanto para las conexiones de entrada ($\gamma = 2,06$) como para las de salida ($\gamma = 3,10$) y las totales ($\gamma = 2,24$) (Pazos *et al*, 2003).

Representando en una gráfica de ejes logarítmicos la conectividad respecto al número de nodos del sistema que muestran dicho comportamiento, he obtenido que la red de reacciones de biodegradación muestra una estructura libre de escala tanto para sus conexiones de entrada y conexiones de salida, situándose su exponente dentro del rango preestablecido para los sistemas biológicos, $\gamma \sim 2,55$ para la conectividad de entrada, $\gamma \sim 2,7$ para la de salida y $\gamma \sim 2,18$ para las conexiones totales (Figuras 20 - 22).

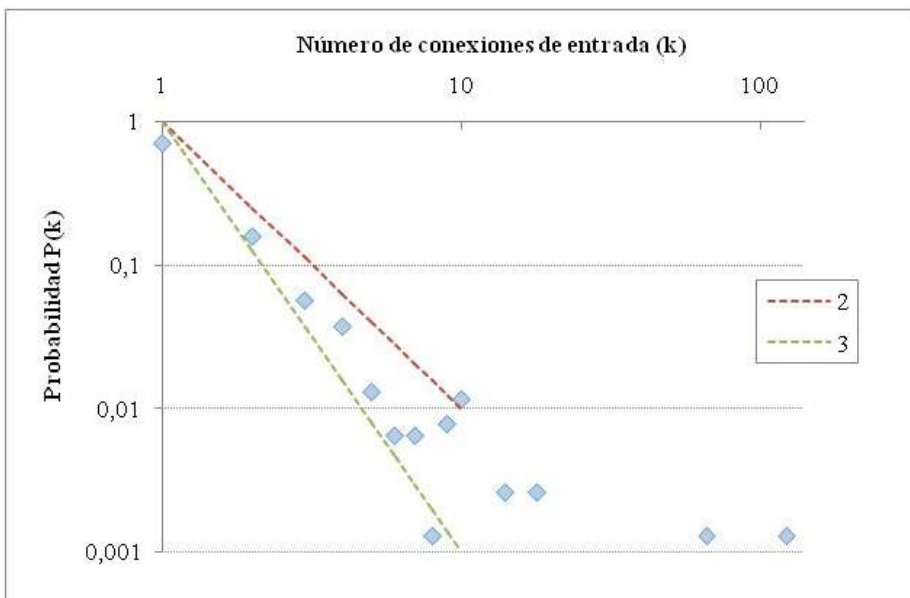


Figura 20. Distribución libre de escala de la conectividad de entrada. Las líneas discontinuas muestran los límites entre los cuales las mayoría de las redes biológicas suelen tener su distribución (la roja para un exponente igual a 2 y la verde para un exponente de 3).

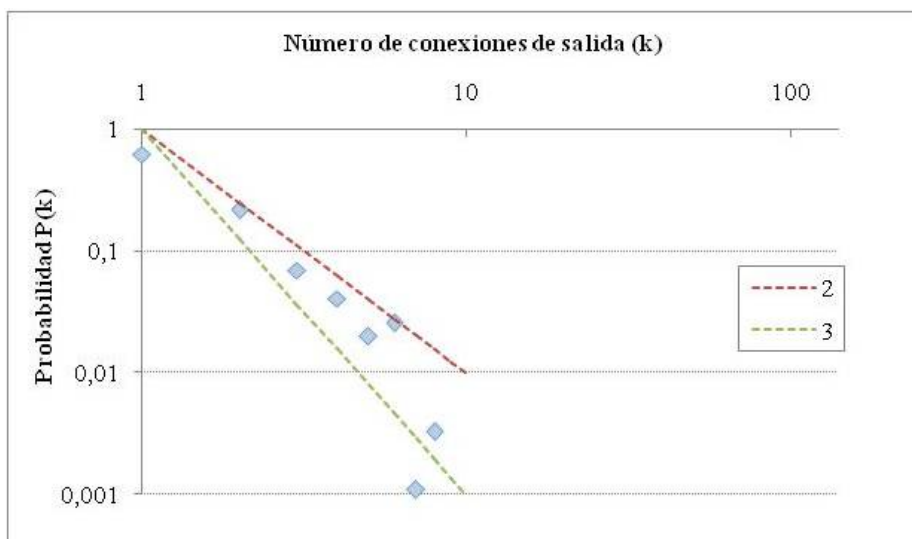


Figura 21. Distribución libre de escala de la conectividad de salida. Las líneas discontinuas muestran los límites entre los cuales la mayoría de las redes biológicas suelen tener su distribución (la roja para un exponente igual a 2 y la verde para un exponente de 3).

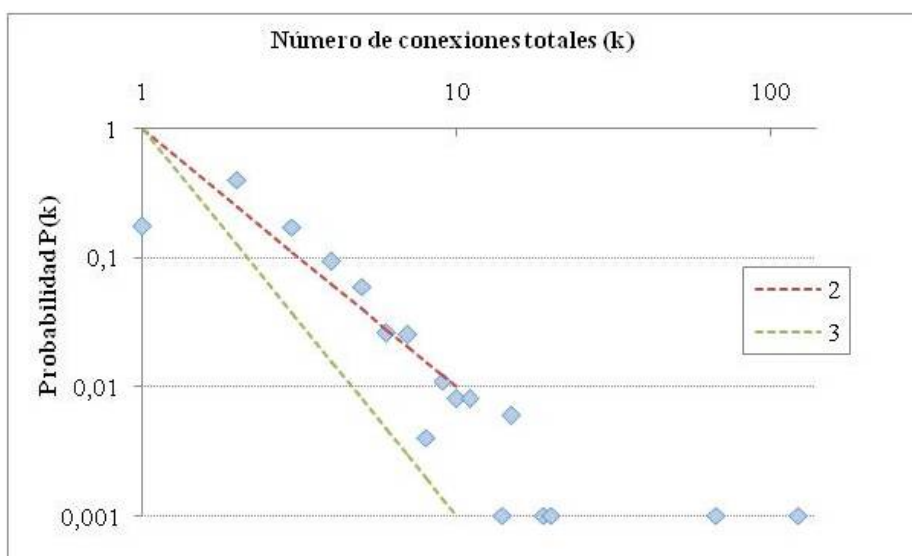


Figura 22. Distribución libre de escala de la conectividad total. Las líneas discontinuas muestran los límites entre los cuales la mayoría de las redes biológicas suelen tener su distribución (la roja para un exponente igual a 2 y la verde para un exponente de 3).

Mis resultados indican que en la red de reacciones de biodegradación es libre de escala y por lo tanto, mientras que la mayoría de sus nodos tienen una conectividad baja, existen nodos altamente conectados a los que se les denomina *hubs*.

4.4.2 Red de “mundo pequeño”.

La segunda característica topológica relevante en el estudio de las redes metabólicas es la relación entre el número de nodos, el diámetro de la red y su coeficiente de *clustering*. Se define diámetro de una red como el tamaño medio de los caminos mínimos entre todos los pares de nodos. A su vez, el coeficiente de *clustering* ó transitividad es la probabilidad de que dos nodos conectados directamente a un tercer nodo, estén conectados entre sí.

Se ha visto que las redes metabólicas tienen una estructura de *Ultra Small World* o Mundo muy Pequeño. La propiedad de mundo pequeño es característica de las redes aleatorias, sin embargo una red *Ultra Small World* es aquella que siendo dispersa, está mucho más clusterizada que una red aleatoria igual de dispersa que ella (con el mismo número de conexiones y grado de conectividad) y tiene un diámetro menor que el de la red aleatoria mencionada. El fenómeno *Small World*, o Mundo Pequeño, indica que la distancia media entre cualquier par de nodos no conectados directamente es pequeña. Esta propiedad dota a las redes de alta eficiencia en la transmisión de información o en el caso de las redes metabólicas para mantener el equilibrio del sistema rápidamente ante la modificación de la concentración de alguno de sus metabolitos.

El diámetro de la red de compuestos de biodegradación es de 5,5 (Pazos *et al*, 2003), algo superior a los encontrados en el estudio de las red de compuestos del Metabolismo Central para 43 organismos, cuyos valores se encontraban entre 2 y 5 (Jeong *et al*, 2000). El estudio de la red del metabolismo central de *Escherichia coli*, tanto para su red de reacciones como para su red de compuestos (Wagner & Fell, 2001), muestra que ambas son redes *Small World*, con un coeficiente de clustering entorno al 0,59 y 0,32, respectivamente y un diámetro de 2,62 y 3,88, respectivamente.

Al comparar la red de reacciones de biodegradación con una red aleatoria de las mismas características (con el mismo número de nodos

y grado de conectividad), he observado que el coeficiente de *clustering* de la red de biodegradación es superior al de la red aleatoria, [ver tabla a continuación], cumpliéndose la primera de las condiciones para ser de “mundo pequeño”, y que el diámetro de la red de biodegradación también es menor que el de la mencionada red aleatoria. Esto parece indicar que es una red de “mundo ultra-pequeño”.

	Red de reacciones de Biodegradación	Red aleatoria análoga
Coef. Clustering Entrada	$7,25 \cdot 10^{-5}$	$3,97 \cdot 10^{-5}$
Coef. Clustering Salida	0,019	$5,58 \cdot 10^{-4}$
Coef. Clustering Total	0,0024	0,0013
Diámetro	8,86	14,19

De nuestro estudio podemos concluir que la red de reacciones de compuestos tóxicos (la red global de biodegradación) presenta una topología “de mundo ultra-pequeño” al igual que lo descrito para otras redes metabólicas como son la red de compuestos y la red de reacciones del Metabolismo Central de *Escherichia coli* (Wagner & Fell, 2001) y la red de compuestos de biodegradación (Pazos *et al*, 2003).

4.4.3 Jerarquía de la red.

La tercera propiedad topológica ampliamente estudiada en las redes metabólicas se relaciona con la existencia de módulos dentro de la red. Los módulos se definen como conjuntos de nodos muy interconectados. Cuando las redes presentan una elevada modularidad, los módulos suelen organizarse jerárquicamente, con una baja interconexión entre ellos, dando lugar a una rápida transición entre puntos.

El primer estudio que introdujo el concepto de jerarquía en el contexto metabólico fue realizado por Ravasz *et al.* en el 2002, al que me he referido anteriormente. En él se describe como las redes del Metabolismo Central de 43 organismos distintos están organizadas en pequeños módulos altamente conectados que se combinan de una manera jerárquica en una organización mayor, menos cohesiva, siguiendo su coeficiente de clustering una ley de potencias (Ravasz *et al*, 2002).

Del mismo modo, representando como se distribuye el coeficiente de *clustering* medio respecto al número de conexiones de los nodos de la red de biodegradación (la red objeto de estudio de esta tesis) he observado que ni el coeficiente de *clustering* respecto a las conexiones de entrada, o a las conexiones de salida o a las totales siguen una ley de potencias. Por lo tanto la red de biodegradación no tiene una estructura jerárquica (Figura 23).

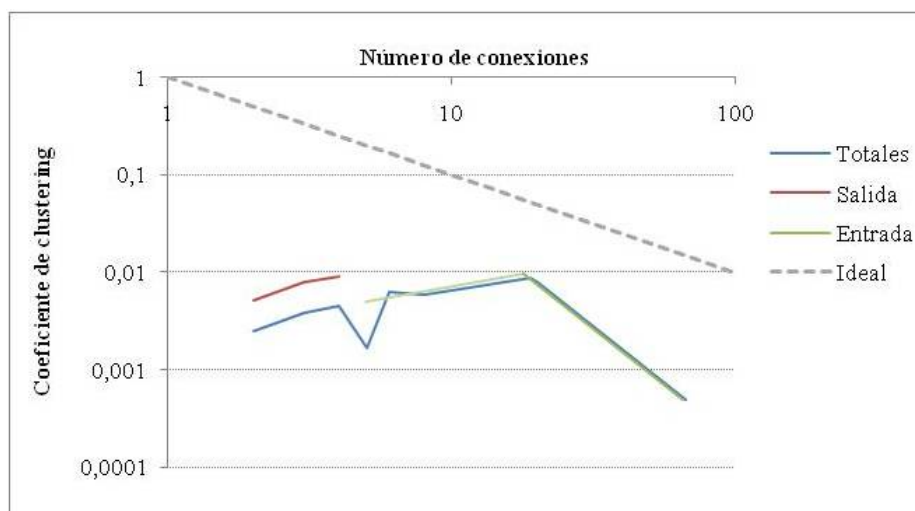


Figura 23. Coeficiente de Clustering en función del número de conexiones de las reacciones. La línea azul representa el comportamiento del coeficiente de clustering teniendo en cuenta el número de conexiones totales, la roja respecto al número de conexiones de salida y la verde en función del número de conexiones de entrada de las reacciones. La línea gris discontinua representa el comportamiento ideal de una red jerárquica.

La relación entre el número de conexiones de entrada y de salida de los nodos, aporta nuevos indicios acerca de la estructura de la red. En los trabajos realizados para la red de compuestos de biodegradación por Pazos en 2003 se observó que presenta una estructura de embudo donde no existen nodos con una alta conectividad de entrada y de salida, a diferencia de lo observado para la red del Metabolismo Central donde existen nodos como el piruvato el cuál es tanto sustrato como producto de un gran número de reacciones. También vió que la conectividad de los compuestos en biodegradación disminuye según se alejan del Metabolismo Central (Pazos *et al*, 2003).

Representando en una gráfica la conectividad de entrada frente a la de salida para cada nodo de la red, puedo dividir el espacio de resultados de acuerdo a la relación que existe entre estos dos tipos de conexiones.

Cuando la conectividad de entrada es mayor que la de salida, los nodos actúan como “embudos”, concentrando el flujo de la red. Sin embargo cuando la conectividad de salida es mayor que la de entrada, los nodos hacen que la red se disperse. Si en cambio el número de conexiones de entrada es igual al número de conexiones de salida, los nodos tienden a mantener lineal la estructura de la red (ni concentran ni dispersan). Por tanto, al representar en una grafica los dos tipos de conectividades (Figura 24) he obtenido que en la red de biodegradación aunque el número de nodos que actúan como embudos es ligeramente mayor, la tónica general es que las reacciones mantienen el flujo hacia el Metabolismo Central, sin concentrarlo o dispersarlo, ya que el par más común es una entrada y una salida.

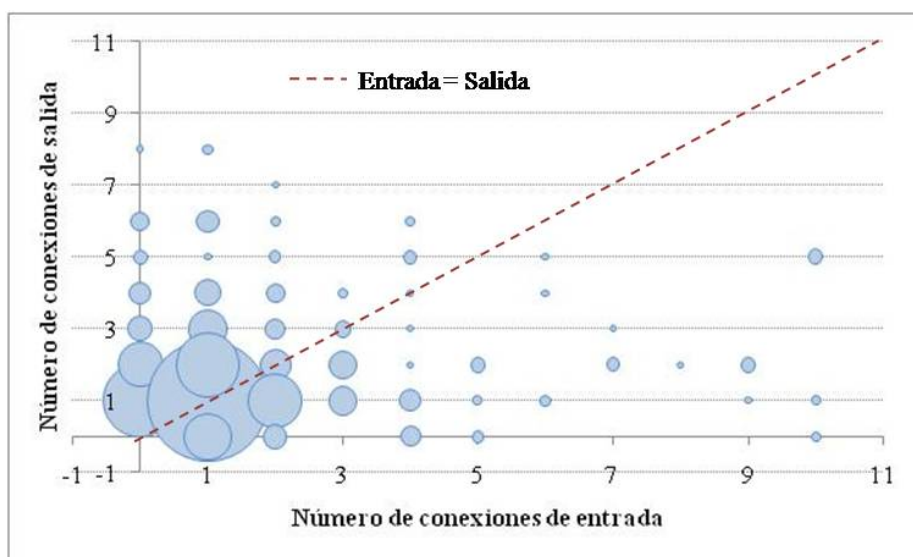


Figura 24. Relación entre el número de conexiones de entrada y salida de las reacciones de la red de biodegradación.

Para estudiar si la distancia al Metabolismo Central influye en la conectividad de las reacciones, he representado la variación de cada tipo de conectividad en función de la distancia (Figura 25 - 27). Los resultados muestran que mientras que la mayoría de las reacciones solo tienen una o dos conexiones de entrada y una o dos conexiones de salida independientemente de su distancia al Metabolismo Central, existe una tendencia a que las reacciones con mayor conexión de entrada, salida o total se sitúen cerca del metabolismo central y que la conectividad decrezca con la distancia.

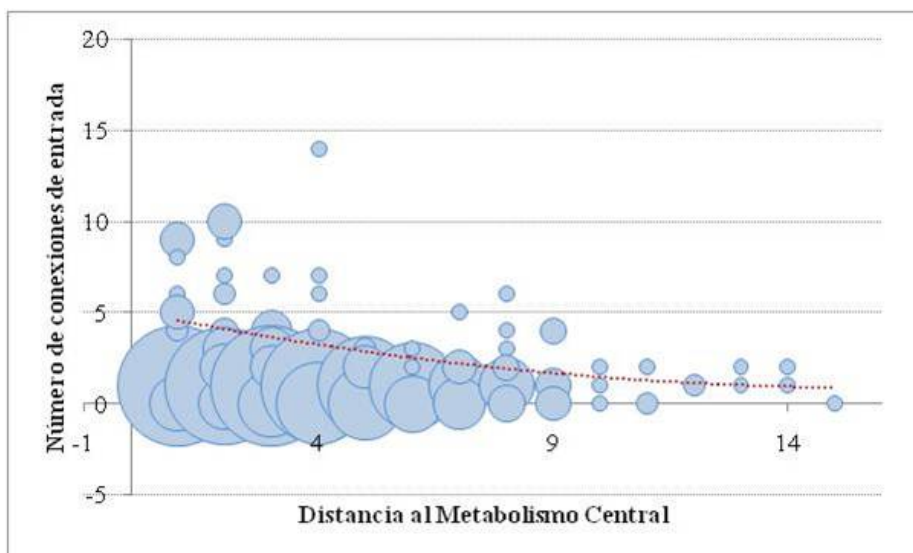


Figura 25. Distribución del número de conexiones de entrada respecto a la distancia de la reacción a la entrada al Metabolismo Central. La línea roja discontinua indica la tendencia.

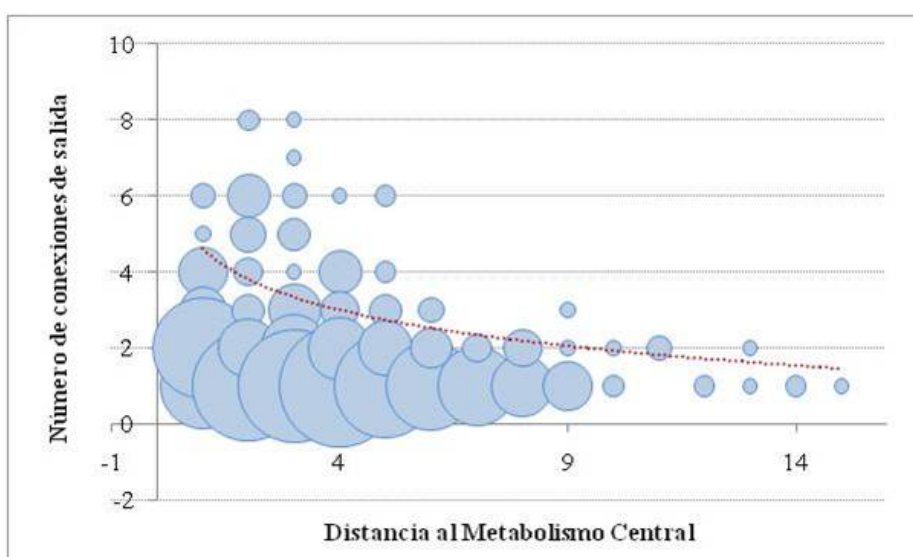


Figura 26. Distribución del número de conexiones de salida respecto a la distancia de la reacción a la entrada al Metabolismo Central. La línea roja discontinua indica la tendencia.

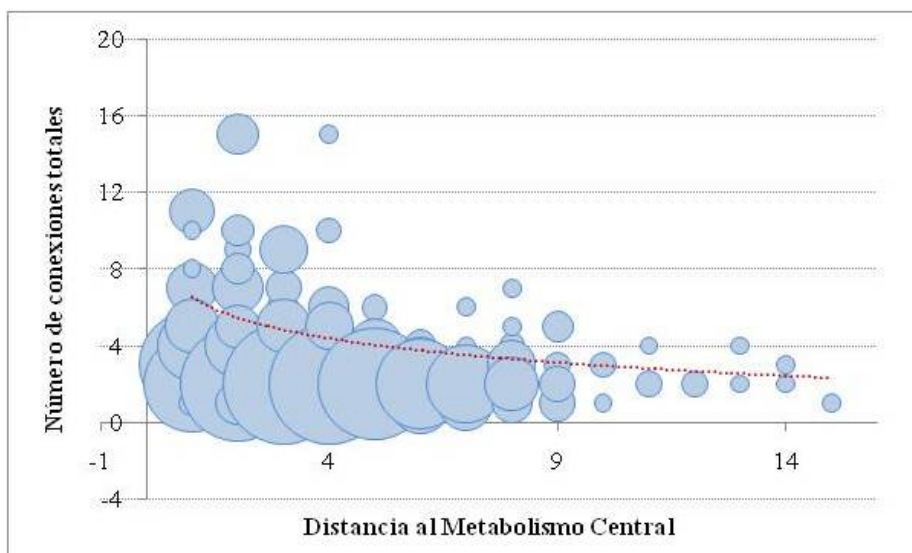


Figura 27. Distribución del número de conexiones totales respecto a la distancia de la reacción a la entrada al Metabolismo Central. La línea roja discontinua indica la tendencia.

Relacionando el número de conexiones con el coeficiente de *clustering*, he encontrado que no existen nodos con una elevada conectividad y coeficiente de *clustering* al mismo tiempo. Esto indica que no hay candidatos a ser centro de clústeres de gran tamaño y que por lo tanto en la red de biodegradación, desde el punto de vista de reacciones, no existen grandes clústeres.

Por lo tanto, a diferencia del comportamiento jerárquico observado en la red del Metabolismo Central, la red de reacciones de biodegradación no presenta jerarquía alguna. Es más, al igual que lo encontrado para la red de compuestos de biodegradación, la red de reacciones de biodegradación presenta una estructura muy lineal, introduciendo cierta interconexión a distancias próximas a la entrada al Metabolismo Central.

4.5 Organización y propiedades de las reacciones en la red de Biodegradación.

Una vez conocida la estructura topológica de la red de reacciones de biodegradación, el siguiente paso va a estar dirigido al estudio de la organización de las reacciones dentro de esta red. Gracias a la asignación de la metaenzima a las reacciones que realizan, podemos estudiar cómo estas metaenzimas se organizan en la red de reacciones de biodegradación. Este análisis revelaría de qué manera la distancia de una reacción a la entrada del Metabolismo Central influye en sus propiedades.

Al haber definido metaenzima como una abstracción de los complejos de proteínas capaces de realizar una reacción y además he encontrado que para todas las reacciones, excepto en dos casos, los complejos de proteínas que tienen asociados una reacción son homólogos, se puede hablar de propiedades de las reacciones como una extrapolación de las propiedades que tiene cualquiera de sus complejos de proteínas.

4.5.1 Complejidad de las metaenzimas encargadas de las reacciones.

La complejidad de una reacción puede definirse como el número de dominios funcionales diferentes asociados a esa reacción que he encontrado en Pfam. Ya que los dominios pueden ser considerados como módulos funcionales dedicados a sub-funciones específicas, la complejidad puede indicar la facilidad con la que las metaenzimas podrían haber tanto aparecido y/o evolucionado a una nueva función, como haberse transferido, ya que se ha demostrado que la HGT ha jugado un papel importante en biodegradación.

He obtenido que en la red de biodegradación, las reacciones tienen asociado entre uno y diez dominios funcionales distintos. La mayoría de las reacciones solo cuentan con uno o dos dominios funcionales. Alrededor de un 60% de las reacciones con proteínas tienen solo un dominio asociado, y solo el 20% de las reacciones tienen más de dos dominios.

Al analizar la complejidad de las reacciones en función de la distancia al Metabolismo Central, encuentro que el 60% y el 75% de las reacciones a distancia uno y dos o tres, respectivamente, solo tienen un dominio funcional, mientras que este valor disminuye hasta prácticamente el 40% a distancias mayores de cinco. En media, las reacciones cercanas al metabolismo tienen 1,5 dominios mientras que en la periferia de la red las reacciones tienen 2,6 dominios (Figura 28).

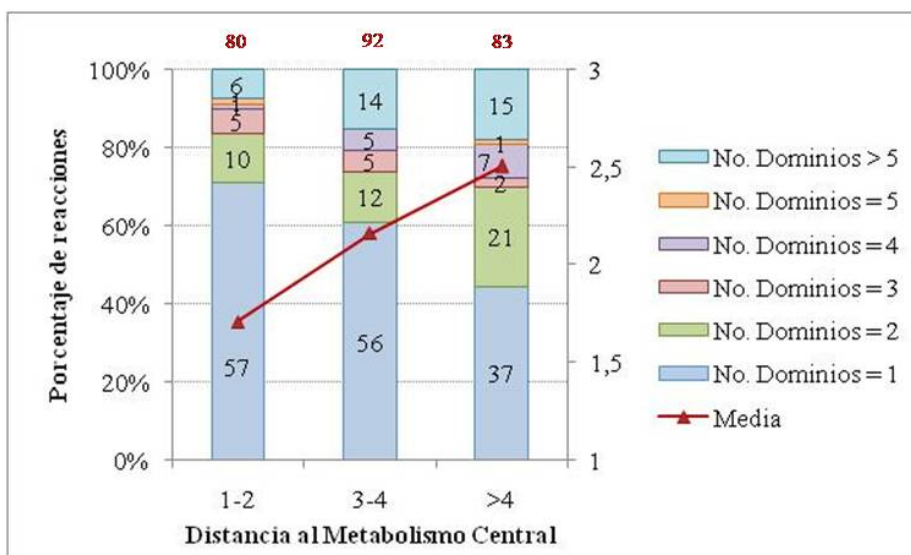


Figura 28. Distribución de la complejidad de las reacciones respecto a la distancia al Metabolismo Central. En la parte superior de la gráfica, las cifras en rojo representan el número total de reacciones con dominios funcionales a cada distancia. Este número de reacciones no incluye a las reacciones sin metaenzima ni las reacciones no conectadas al CM. Los valores sobre la figura son número de casos, no porcentajes.

La complejidad medida como el número de dominios asociados a una cierta reacción, podría deberse a dos fenómenos diferentes: al aumento en el número de dominios por proteína o al aumento en el número de subunidades por complejo. Para saber cuál de estos dos fenómenos es más habitual en las reacciones de biodegradación, he calculado la distribución respecto a la distancia al metabolismo central tanto del número de dominios por proteína, como del número de proteínas por complejo.

Referente al número de dominios por proteína, como máximo en una proteína he encontrado tres dominios funcionales. El 20% de las proteínas a uno o dos pasos del Metabolismo Central tienen más de un

dominio, mientras que a distancias mayores de cuatro pasos este número asciende al 40% de las proteínas (Figura 29).

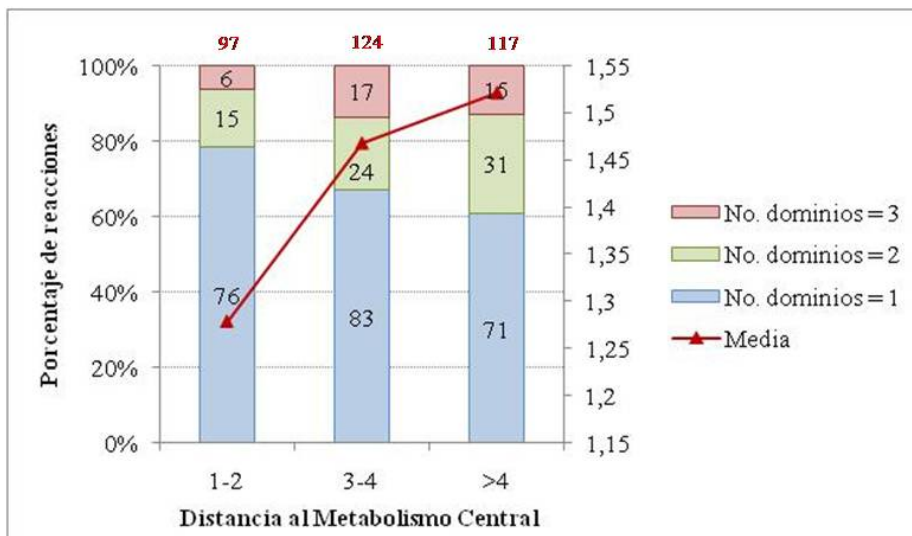


Figura 29. Distribución respecto a la distancia a la entrada al Metabolismo Central del número de dominios por proteína. En la parte superior de la gráfica, las cifras en rojo representan el número total de proteínas con dominios funcionales a cada distancia, teniendo en cuenta que si una proteína forma parte de un complejo que participa en más de una reacción, aparecerá en todas las distancias de esas reacciones y tantas veces como reacciones esté asociada. Los valores sobre la figura son número de casos, no porcentajes.

Respecto al número de proteínas por complejo, la tendencia es similar. El porcentaje de complejos con una sola proteína desciende un 10% conforme que la distancia aumenta. Además a más de tres pasos del metabolismo central el porcentaje de complejos con tres o más proteínas (un 30% y un 40%, a distancias tres y cuatro y mayor de cuatro respectivamente) es superior al encontrado cerca del CM (alrededor de un 10%) (Figura 30).

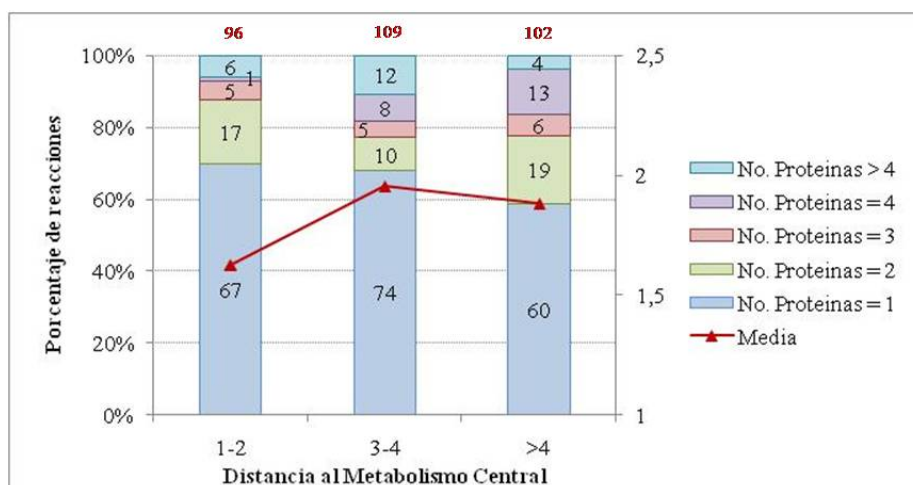


Figura 30. Distribución del número de subunidades por complejo de proteínas respecto a la distancia a la entrada al Metabolismo Central. En la parte superior de la gráfica, las cifras en rojo representan el número total de complejos a cada distancia, teniendo en cuenta que si un complejo aparece en más de una reacción, se contabiliza tantas veces como reacciones participe y a las distancias en las que se encuentren esas reacciones. Los valores sobre la figura son número de casos, no porcentajes.

La ruta de biodegradación del 3-fenilpropionato (Diaz *et al*, 1998) ejemplifica este comportamiento. Mientras que el primer paso de su degradación es realizado por un complejo formado por 4 proteínas, donde una de ellas tiene dos dominios funcionales (siendo tres el máximo), el resto de las reacciones, más cercanas al CM, solo tienen complejos formados por una proteína con un solo dominio funcional (Figura 31).

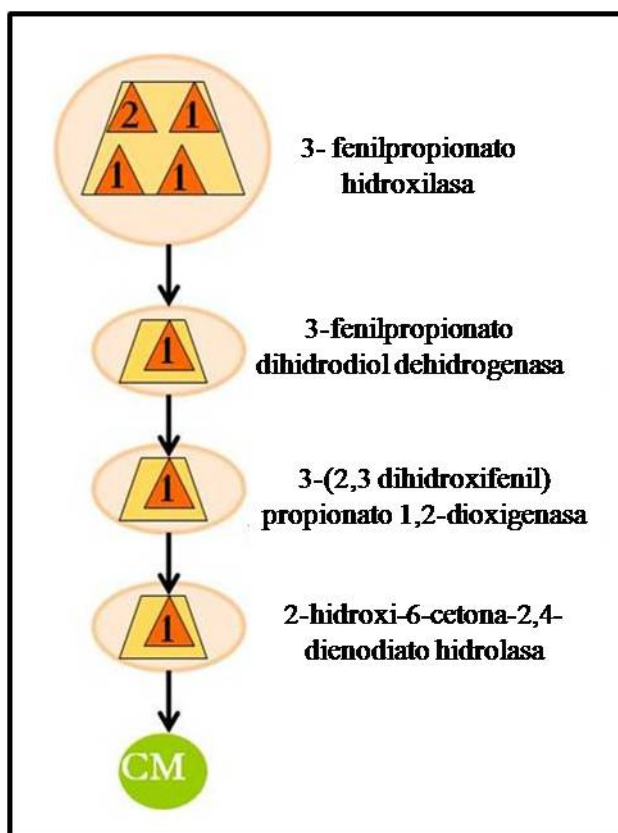


Figura 31. Ejemplo complejidad en la red de Biodegradación.

En resumen, en la red de biodegradación la complejidad aumenta a medida que nos alejamos del Metabolismo Central. Ésta es debida tanto al incremento del número de dominios por proteína como al aumento del número de proteínas que tienen los complejos.

4.5.2 Especificidad de las metaenzimas que realizan las reacciones.

La especificidad de una reacción se define como el número de reacciones distintas que es capaz de realizar su complejo, y hace referencia al grado de adaptación del complejo a su función (o reacción). Complejos muy específicos están asociados con una y solo una reacción en la red de biodegradación, mientras que complejos más promiscuos están relacionados a más de una reacción.

La mayoría de los complejos de proteínas, el 70%, se encontraron asociados con solo una reacción, aunque existen complejos capaces de realizar hasta nueve reacciones diferentes.

En cuanto a su distribución en la red, más del 60% de las reacciones a uno o dos pasos del metabolismo tienen asociado un complejo, mientras que para distancias mayores este valor se reduce al 40%. En zonas alejadas del metabolismo (a más de cuatro pasos) el porcentaje de reacciones con complejos que realizan cuatro o más reacciones asciende al 31% mientras que en las cercanías del metabolismo es de un 7% (Figura 32).

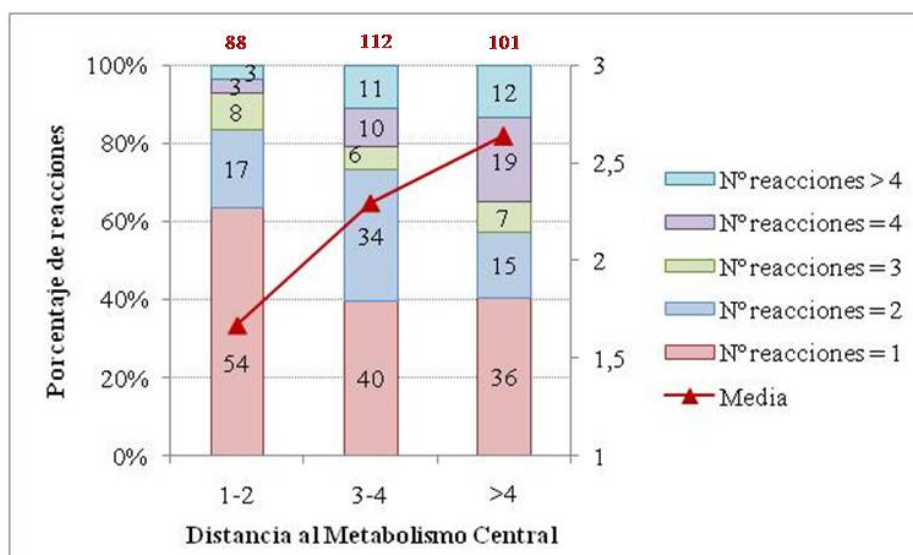


Figura 32. Distribución de la especificidad de los complejos de las reacciones en biodegradación respecto a la distancia a la entrada al Metabolismo Central. En la parte superior de la gráfica, las cifras en rojo representan el número total de reacciones con un complejo que está asociado a un número de reacciones. Un complejo se contabiliza tantas veces como reacciones participe. Los valores sobre la figura son número de casos, no porcentajes.

Un ejemplo de este comportamiento es el complejo de proteínas de *Pseudomonas sp.* P51 codificado en los genes: tcbAa, tcbAb, tcbAc y tcbAd, que funciona como una dioxigenasa clorobenzénica, ejemplifica el comportamiento anterior. Este complejo enzimático es capaz de realizar un gran número de reacciones diferentes, hidroxilando (añadiendo grupos OH) compuestos con un anillo en su estructura. Aunque el substrato principal de esta proteína es el clorobenceno,

puede degradar de manera secundaria otros muchos substratos cuya estructura es similar, como el tolueno, el naftaleno o el bifenilo. De esta forma, el complejo presenta poca especificidad a substrato apareciendo en los primeros pasos, en la periferia de la red, de muchas rutas de biodegradación (Figura 33) (Werlen *et al*, 1996).

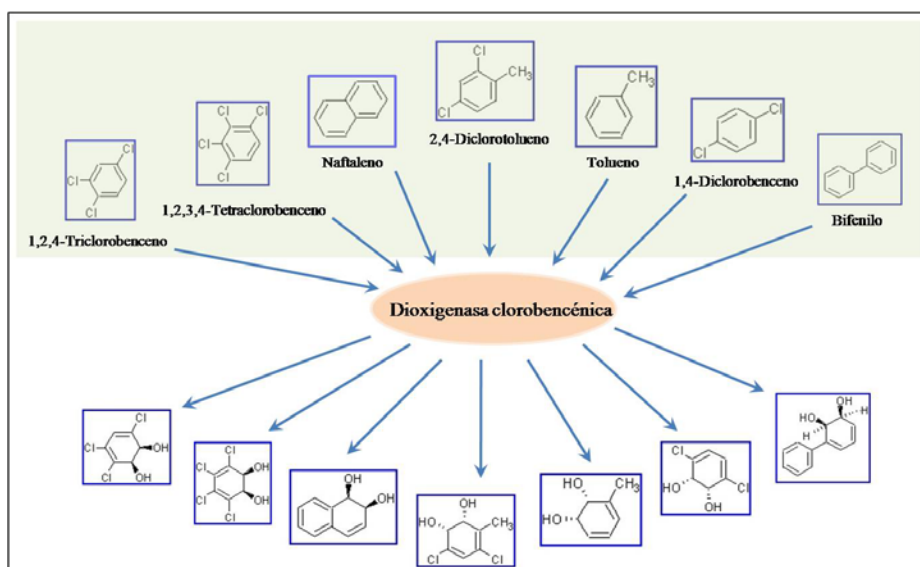


Figura 33. Ejemplo especificidad en la red de biodegradación.

Por lo tanto en la red de biodegradación, las reacciones suelen ser específicas. Sin embargo su distribución en la red no es homogénea, perdiéndose la especificidad volviéndose más promiscuas a medida que nos alejamos del metabolismo central.

4.5.3 Variabilidad de las proteínas.

La variabilidad de una proteína se define como la velocidad o el grado de libertad a la que evoluciona. Para calcularlo he buscado las proteínas homólogas por identidad de secuencia de cada una de las proteínas de la red de biodegradación en las bases de datos de secuencias Swissprot y Trembl.

Estudios realizados en relación a la transferencia de función entre proteínas (Devos & Valencia, 2000; Devos & Valencia, 2001; Tian & Skolnick, 2003; Hatzimanikatis *et al*, 2004) han postulado que si el porcentaje de identidad entre dos proteínas es superior al 40%, se puede

realizar transferencia de la actividad catalítica entre ellas, pero si se quiere mantener la especificidad de sustrato, el porcentaje de identidad debe ser superior al 60%. Por tanto modificando el porcentaje de identidad a partir del cual extraigo de las bases de datos de secuencias los homólogos de las proteínas de biodegradación, controlo si los homólogos conservarían la especificidad a sustrato y la actividad catalítica.

Teniendo en cuenta lo anterior, en este experimento trato de averiguar la relación entre la localización de las proteínas dentro de la red de biodegradación y el grado de conservación de las secuencias de las mismas. Los porcentajes de identidad que he utilizado son del 95%, con el que es bastante seguro que la especificidad a sustrato se mantenga y del 50% con el cual la especificidad a sustrato se ha perdido pero la conservación de la actividad catalítica es bastante segura.

Cuando el porcentaje de identidad es del 95% el número de homólogos por proteína aumenta con la distancia. De esta forma mientras que el solo alrededor del 15% de las proteínas a distancia uno o dos tienen cuatro o más proteínas homólogas en Swissprot y Trembl, este porcentaje se dobla ascendiendo al 30% de las proteínas a distancias superiores de cuatro pasos del CM (Figura 34).

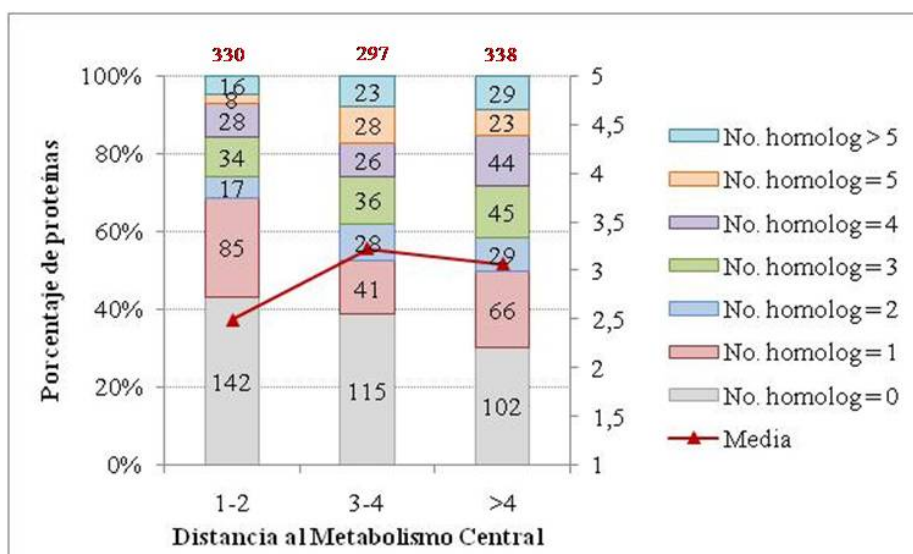


Figura 34. Distribución del número de homólogos de las proteínas de la red de biodegradación, respecto a su distancia a la entrada del Metabolismo Central, utilizando un porcentaje de identidad entre secuencias homólogas superior al 95%. En la parte superior de la gráfica, las cifras en rojo representan el número total

de proteínas, teniendo en cuenta que si una proteína forma parte de un complejo que realiza varias reacciones, la proteína aparecerá tantas veces como reacciones participe. Los valores sobre la figura son número de casos, no porcentajes.

Finalmente, bajando el porcentaje de identidad al 50%, se observa como el número de homólogos por proteína de biodegradación se vuelve constante e independiente de su localización dentro de la red. Ahora a lo largo de toda la red, alrededor del 60% de las proteínas tienen cuatro o más homólogos (Figura 35).

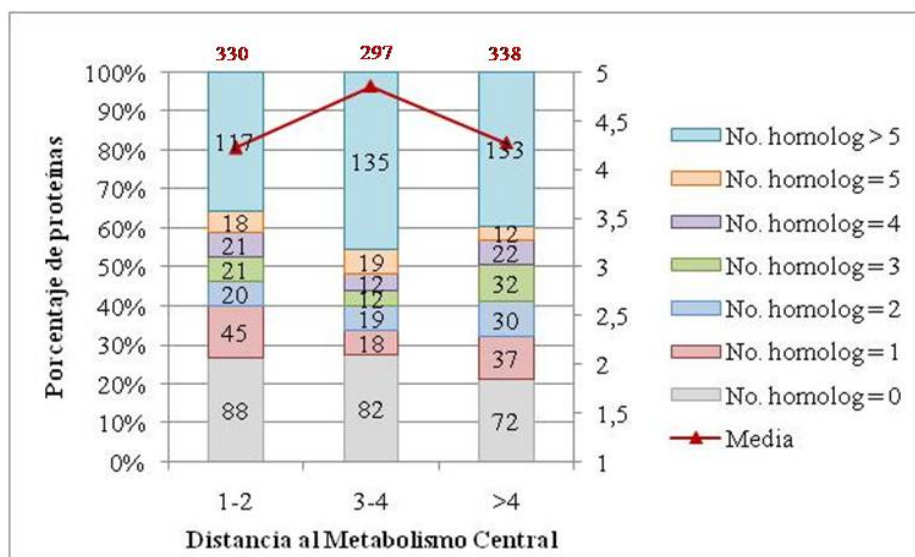


Figura 35. Distribución del número de homólogos de las proteínas de la red de biodegradación, respecto a su distancia a la entrada del Metabolismo Central, utilizando un porcentaje de identidad entre secuencias homólogas superior al 50%. En la parte superior de la gráfica, las cifras en rojo representan el número total de proteínas, teniendo en cuenta que si una proteína forma parte de un complejo que realiza varias reacciones, la proteína aparecerá tantas veces como reacciones participe. Los valores sobre la figura son número de casos, no porcentajes.

Los resultados obtenidos muestran que cuando se usa un umbral que agrupa proteínas que tendrían la misma especificidad (95%), las proteínas alejadas del Metabolismo Central tienen un mayor número de homólogos que las proteínas próximas al Metabolismo Central. Sin embargo cuando se usa un umbral que no garantiza la especificidad aunque si la actividad catalítica (50%), las proteínas a lo largo de toda la red presentan un número similar de homólogos.

Como resumen de este apartado, la complejidad de las reacciones en la red de reacciones de biodegradación aumenta con la distancia, al

mismo tiempo que disminuye su especificidad (son más promiscuas en la periferia de la red). Además, los grupos de proteínas homólogos alejados del Metabolismo Central tienen menos diversidad de secuencia (las proteínas se parecen más entre sí) que los que están cerca.

4.6 Formación y evolución de la red de Biodegradación.

Los resultados obtenidos a cerca de la organización de las reacciones dentro de la red de biodegradación y sus implicaciones en la adaptación, evolución y dispersión de las actividades, han ofrecido una primera aproximación hacia los mecanismos de evolución que subyacen en la red. Mientras que ya han sido realizados estudios sobre la posible formación de rutas individuales en biodegradación (van der Meer *et al*, 1998; Arengi *et al*, 2001; Johnson & Spain, 2003; Shapir *et al*, 2007), en este apartado expongo el primer estudio sobre la posible formación y evolución de la red global de reacciones de biodegradación.

Las perturbaciones humanas a gran escala en el medio ambiente son relativamente recientes, prácticamente desde la revolución industrial (Diaz, 2004). Debido a que la mayoría de estos compuestos tóxicos no son compuestos que de manera natural se encuentren en el medio, es probable que las rutas de biodegradación de estos productos hayan sido ensambladas recientemente, en términos evolutivos. Este hecho ha provocado que los microorganismos desarrollen nuevas capacidades metabólicas para hacer frente a esta nueva situación (Barkay & Pritchard, 1988; Bouwer & Zehnder, 1993; Liu & Suflita, 1993; Springael & Top, 2004). Debido a esto, la red de biodegradación es un modelo óptimo para estudiar las primeras etapas en la formación y evolución de una nueva red metabólica.

Es factible pensar que si los microorganismos han tenido que adaptarse rápidamente a la reciente aparición de grandes cantidades de compuestos tóxicos en el medio, las proteínas de las rutas de biodegradación deberían de provenir de alguna proteína ya existente antes de la actividad industrial, las cuales habrían sufrido mutaciones recientes y por lo tanto serían cercanas en términos evolutivos. Sin embargo, de manera alternativa la tasa de mutación podría estar acelerada y por ello existir una alta divergencia entre las proteínas de biodegradación y sus precursores.

4.6.1 Comparación funcional entre Biodegradación y el Metabolismo Central.

Ya que el Metabolismo Central pre-existe a la introducción de compuestos recalcitrantes, puede que las enzimas que realizan procesos de biodegradación estuvieran relacionadas evolutivamente con enzimas del Metabolismo Central. Por ello, voy a comenzar el estudio de la evolución utilizando los dominios funcionales presentes en biodegradación, y analizando su presencia y frecuencia en el Metabolismo Central para cuantificar las diferencias funcionales existentes entre ambas redes.

Analizando para cada dominio de degradación, el número de reacciones en los que se encuentra, he obtenido un ranking en el que se indica su abundancia en la red [ver tabla Apéndice 2]. El dominio más frecuente en las reacciones de biodegradación es el relacionado con el transporte de electrones “Dominio de unión al grupo 2Fe-2S” (*2Fe-2S iron-sulfur cluster binding domain*, PF00111), apareciendo en un 5,29% de las reacciones de biodegradación con descripción funcional. A continuación se sitúan familias relacionadas con actividades red-ox como son el dominio “*Rieske [2Fe-2S]*” (PF00355), el cual aparece un 4,55%, y los dominios de oxidoreductasas de unión a FAD (PF00970) y de unión a NAD (PF00175), cada uno de los cuales aparece un 4,44%. Por último, dentro de los dominios más frecuentes se encuentran las subunidades alfa y beta del “Anillo aromático de hidroxilación dioxigenasa” (*Ring hydroxylating*, PF00848 y PF00866) los cuales aparecen un 2,96% y un 3,38% respectivamente, siendo la primera de las subunidades la unión al ión hierro y la segunda la actividad catalítica del complejo que forman. Estos dos últimos dominios son muy frecuentes en biodegradación debido a su participación en la degradación de compuestos aromáticos por bacterias aeróbicas, convirtiendo estructuras en anillo cerrado a cis-dioles no aromáticos (Neidle *et al*, 1991).

Una vez conocida la frecuencia de las familias funcionales dentro de la red de biodegradación, comparé este valor con su frecuencia en las reacciones del Metabolismo Central [ver tabla Apéndice 2]. Este análisis revela qué dominios presentes en biodegradación están sobre representados en una u otra red. Para realizar el estudio, he representado en una gráfica el porcentaje de reacciones de biodegradación que tienen ese dominio frente al número de veces que

este porcentaje es superior al presentado en el Metabolismo Central. En espacio de resultados obtenido se pueden identificar diversos grupos de dominios en biodegradación (Figura 36).

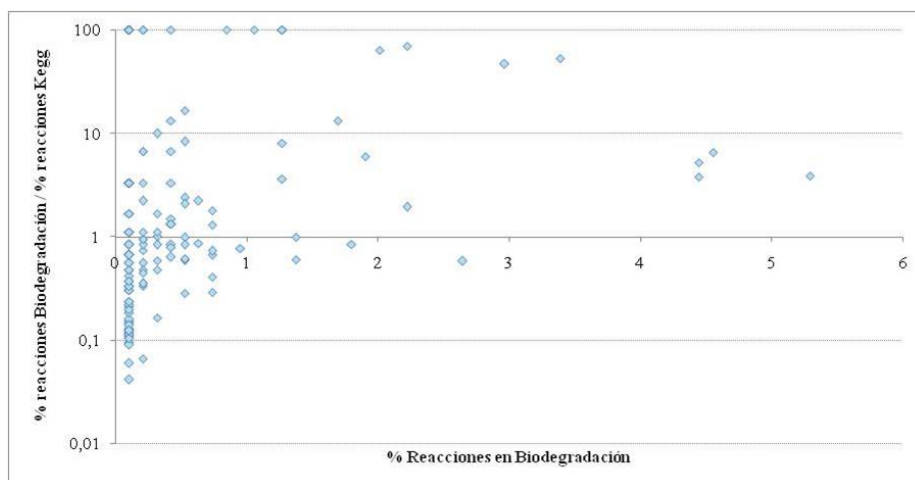


Figura 36. Frecuencia de los dominios funcionales de biodegradación respecto a su frecuencia en la red del Metabolismo Central.

La zona izquierda – inferior de la gráfica agrupa los dominios funcionales que son mucho más abundantes en el Metabolismo Central que en biodegradación. El caso más significativo de este grupo es el del “Dominio oxidoreductasa dependiente de FAD” (PF01266), el cual aparece en solo una reacción de biodegradación mientras que en el metabolismo aparece en 80 reacciones (en un 2,5% de las reacciones del metabolismo central). Se ha descrito que este dominio participa en el metabolismo de los amino ácidos y la respiración.

Existe otro grupo de dominios, situados en la parte izquierda superior, que siendo poco frecuentes en biodegradación (a lo sumo en un 1,26% de las reacciones) son exclusivos de biodegradación, no se encuentran en el metabolismo central. En este grupo encontramos dominios tales como la “Fenol hidroxilasa conservada” (*phenol hydroxylase conserved*, PF04663), la “Subunidad fenol hidroxilasa” (*phenol hydroxylase subunit*, PF06099), la “Catecol dioxigenasa N terminal” (*catechol dioxygenase N terminus*, PF04444), la “Familia dienolactato hidroxilasa” (*Dienelactone hydroxylase family*, PF01738), la “Alquilmercurio liasa” (*Alkylmercury lyase*, PF03243) y la “Enzima tautomerasa” (*Tautomerase enzyme*, PF01361). Todos ellos son dominios que participan en la degradación de compuestos aromáticos,

excepto el dominio “Alquilmercurio liasa” que se ocupa de la eliminación de los enlaces carbón-mercurio en compuestos organomercuriales.

También hay un grupo de dominios en la parte central superior del gráfico, donde se agrupan los dominios que aunque no son exclusivos de biodegradación, son mucho más frecuentes en biodegradación que en el Metabolismo Central. Aquí se encuentran el dominio correspondiente a la subunidad beta del “Anillo aromático de hidroxilación dioxigenasa” (*Ring hydroxylating*) el cual además de aparecer en la degradación de compuestos aromáticos, participa dentro del Metabolismo Central en el metabolismo de la melanina.

Por último a la derecha de la gráfica, se encuentra el grupo de dominios más frecuente en biodegradación, los ya mencionados “Dominio de unión al grupo 2Fe-2S” (*2Fe-2S iron-sulfur cluster binding domain*, PF00111) y relacionados con actividades redox (PF00175, PF00970 y PF00355). Este grupo de funciones son aproximadamente cuatro veces más habituales en biodegradación que en el Metabolismo Central. En el caso de “Dominio de unión al grupo 2Fe-2S”, este comportamiento se debe a que en el Metabolismo Central aparece en el ciclo del ácido cítrico, y en la red de biodegradación en la degradación del metano y ftalato.

Con los resultados obtenidos, he observado que la red de biodegradación está enriquecida con dominios relacionados con actividades de oxido- reducción. De la comparativa entre la frecuencia de los dominios en biodegradación y en la red del Metabolismo Central, se observa que los dominios exclusivos de biodegradación, son poco frecuentes y que los dominios más frecuentes en biodegradación también están presentes en el Metabolismo Central (aunque en frecuencia algo menores).

4.6.2 Evolución de las proteínas de biodegradación.

Terminada la comparación funcional entre las dos redes metabólicas, la de biodegradación y la del Metabolismo Central, continuaré el estudiando de sus relaciones evolutivas identificando sus homólogos mediante comparación de sus secuencias. Para ello comenzaré el análisis buscando los homólogos de las proteínas de biodegradación

dentro de la red del Metabolismo Central. Al analizar las secuencias de las proteínas de la red de biodegradación, identificando homólogos con un porcentaje de identidad superior al 30% (para asegurar que comparten una relación evolutiva común), he obtenido que únicamente alrededor del 25% de las proteínas de biodegradación tienen un homólogo identificable en el Metabolismo Central (Figura 37).

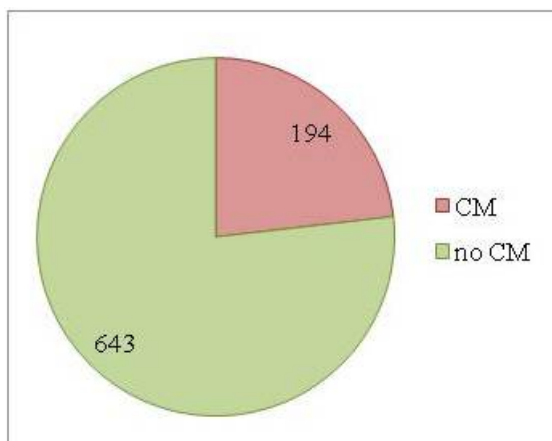


Figura 37. Relación evolutiva detectable mediante homología de secuencias de las proteínas de biodegradación con las del Metabolismo Central.

Para las proteínas que no he tienen una relación evolutiva identificable por comparación de secuencia con el Metabolismo Central, he ampliado su búsqueda en rutas de procesos celulares (movilidad, crecimiento, muerte, comunicación, etc.), procesos de información genética (transcripción, traducción, etc.), enfermedades humanas y procesos de información medioambiental (Ver Apéndice 3 para descripción de las rutas). Sin embargo, tampoco encontré ningún homólogo suyo en estos sistemas. Por último he estudiado si estas proteínas sin similitud de secuencia tienen homólogos cercanos dentro de la propia red de biodegradación, encontrando que el 55% de ellas si tenía.

Los resultados obtenidos en el estudio de la evolución a nivel de proteínas muestran que solo un porcentaje muy pequeño de las proteínas en biodegradación (el 25%) tienen una relación evolutiva identificable con alguna proteína del Metabolismo Central. El 75% de las proteínas que no tienen relación evolutiva detectable con el metabolismo central, tampoco lo tenían en otros sistemas biológicos a

excepción de la propia red de biodegradación, donde el 55% de ellos tenían un homólogo.

4.6.3 Evolución de las reacciones de biodegradación.

Las proteínas no actúan de manera aislada en la red de biodegradación, sino que se asocian formando complejos cuya acción conjunta da lugar a las reacciones. Los diferentes papeles de estas proteínas dentro del complejo pasan desde realizar la actividad catalítica, la unión a sustrato o la unión al cofactor. Por ello el siguiente paso fue estudiar la formación de los complejos. Como ya he mencionado anteriormente, los complejos asociados a una misma reacción (es decir, la metaenzima) son homólogos entre sí, excepto para los dos casos descritos en el apartado 4.1. Por ello, se puede extrapolar las propiedades de las metaenzimas a las reacciones seleccionando para cada reacción el complejo con el mayor número de subunidades, el más completo (ver métodos).

Los resultados que he obtenido muestran que el 93% de las reacciones tienen asociado un complejo homogéneo, es decir, o todas sus proteínas tienen una relación evolutiva identificable (reacciones “CM”) o ninguna de ellas (reacciones “no CM”). El 7% de reacciones restante tiene asociado un complejo en el que algunas de sus proteínas tienen relación evolutiva identificable con el Metabolismo Central y otras no, reacciones “mixtas” (Figura 38).

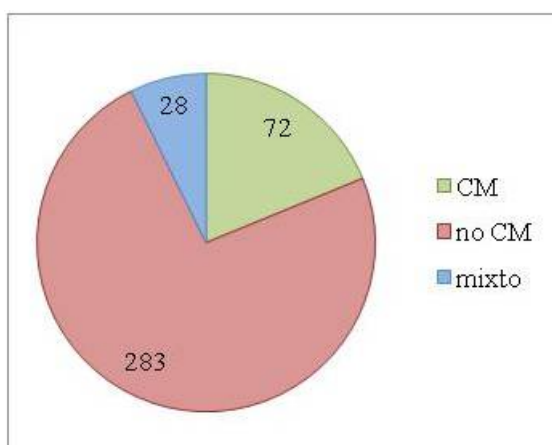


Figura 38. Relación evolutiva detectable mediante homología de secuencia de los complejos de biodegradación en la red del Metabolismo Central.

Para saber si los resultados obtenidos en el análisis anterior son solo fruto de la frecuencia de proteínas “CM” o “no CM” en la red, he repetido el experimento utilizando mil redes aleatorias análogas a la de biodegradación (es decir, con el mismo número de reacciones, y donde cada complejo asociado a la reacción tiene el mismo número de proteínas) pero con la diferencia de que las proteínas que forman cada uno de esos complejos es aleatoria. La media de todas las simulaciones mostró un mayor número de reacciones “mixtas” que las encontradas en la red de biodegradación (siendo el z-score de -24,78). Por el contrario, el número de reacciones “CM” y “no CM” era muy superior en la red de biodegradación que la media de la simulaciones (con un z-score de 87,49 para relación con CM y 19,00 para no relación con CM). Por lo tanto, los resultados obtenidos en la red de biodegradación parecen significativos y no producto de su frecuencia.

Tipo de Reacciones	Media	Desviación Típica
CM	0.302	0.8194
No CM	56.931	11.89
Mixta	325.767	12.014

La conclusión de este apartado es que son muy pocas las reacciones cuyos complejos de proteínas (complejo-enzima) tienen una relación evolutiva cercana con el Metabolismo Central, y que solamente en unos pocos casos las reacciones cuentan con complejos de proteínas mixtos. Analizando más en detalle el 7% de reacciones mixtas he visto que pueden ser de dos tipos, Mientras que el 64% de ellas tienen complejos donde a excepción de una de sus proteínas, ninguna tiene relación evolutiva cercana con el Metabolismo Central, el 36% de reacciones restante tiene un complejo donde a excepción de una de sus proteínas, todas tienen relación evolutiva cercana con el Metabolismo Central.

La distribución de los diferentes tipos de reacciones (“CM”, “no CM” y “mixtas”) a lo largo de la red de biodegradación, muestra que mientras que reacciones con complejos homogéneos (reacciones “CM” y “no CM”) están distribuidos de manera constante en todas las distancias, las reacciones mixtas se encuentran habitualmente alejadas del CM, encontrándose el 90% de ellas a 3 o más pasos del Metabolismo Central (Figura 39).

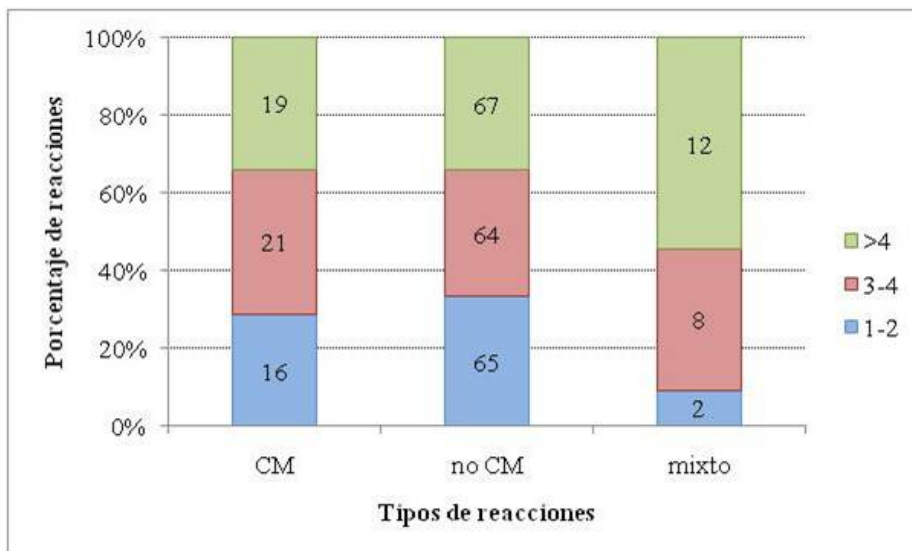


Figura 39. Distribución de los tipos de reacciones (CM, no CM y mixtas) respecto a su distancia a la entrada del Metabolismo Central.

La figura 40 muestra los dos tipos de reacciones mixtas que ejemplifican tanto la distancia a la que se suelen encontrar como las pautas posibles de formación de sus complejos. Por una parte, el complejo de la reacción que degrada el 1,4-diclorobenceno, la cual se realiza a ocho pasos del metabolismo central, tienen todas sus proteínas una relación evolutiva cercana con proteínas del metabolismo de la fenilalanina, a excepción de una proteína que no tiene relación evolutiva cercana con el Metabolismo Central y funcionalmente contiene un dominio Rieske [2Fe-2S] (TcbAd). En este caso parece que la actividad catalítica la adquiere del Metabolismo Central y la proteína sin relación evolutiva cercana añade la unión a cofactor.

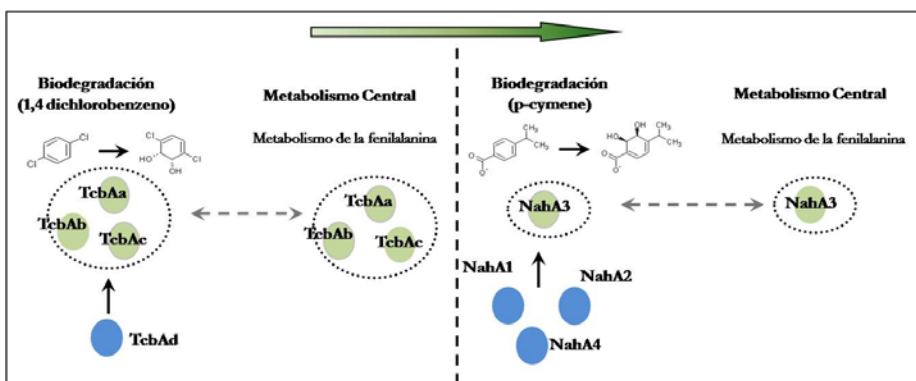


Figura 40. Ejemplos del origen de las reacciones mixtas.

El segundo ejemplo describe la degradación del compuesto p-cinamil alcohol (*p-cymene*), la cual se realiza a cinco pasos del Metabolismo Central (al inicio de la ruta). Aquí, únicamente la actividad catalítica (NahA3, un dominio *Ring hydroxylating alpha subunit*) ha sido adquirido desde el metabolismo de la fenilalanina. El resto de proteínas (NahA1, NahA2 y NahA4) no tienen relación evolutiva con el Metabolismo Central y lo que aportan a la reacción es su unión a cofactor y sustrato.

4.6.4 Evolución del ensamblaje de las reacciones de la red.

Una vez analizadas las relaciones evolutivas de las reacciones, quise saber si dos reacciones que tienen relación evolutiva cercana con el Metabolismo Central y son consecutivas en biodegradación, sus reacciones relacionadas por similitud de secuencia en el Metabolismo Central también son consecutivas. Este análisis mostró que excepto para tres casos, el resto de pares de reacciones consecutivas con proteínas homologas en el Metabolismo Central, sus homólogos en éste eran también consecutivos (Figura 41). Además el 60% de estos bloques se encuentran en las proximidades de la entrada al Metabolismo Central (a uno o dos pasos).

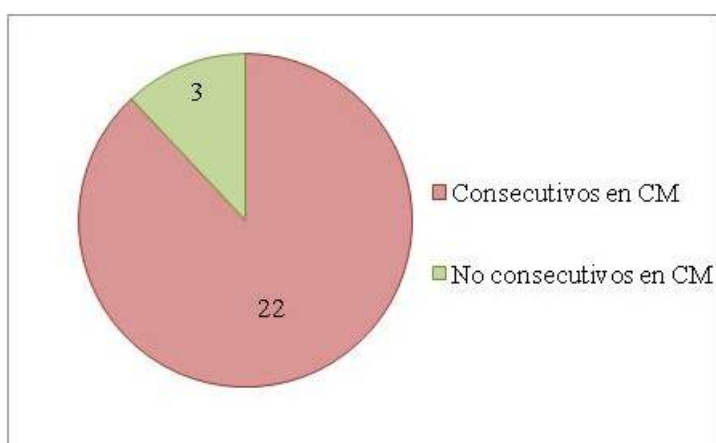


Figura 41. Pares de reacciones consecutivas en biodegradación cuyos homólogos también lo son en la red del Metabolismo Central.

Un caso que ejemplifica lo anterior es el que se muestra en la figura 42 donde tres actividades consecutivas de la red de biodegradación, localizadas en las proximidades del Metabolismo Central, tienen actividades enzimáticas relacionadas por similitud de secuencia en el Metabolismo Central, siendo éstas también consecutivas.

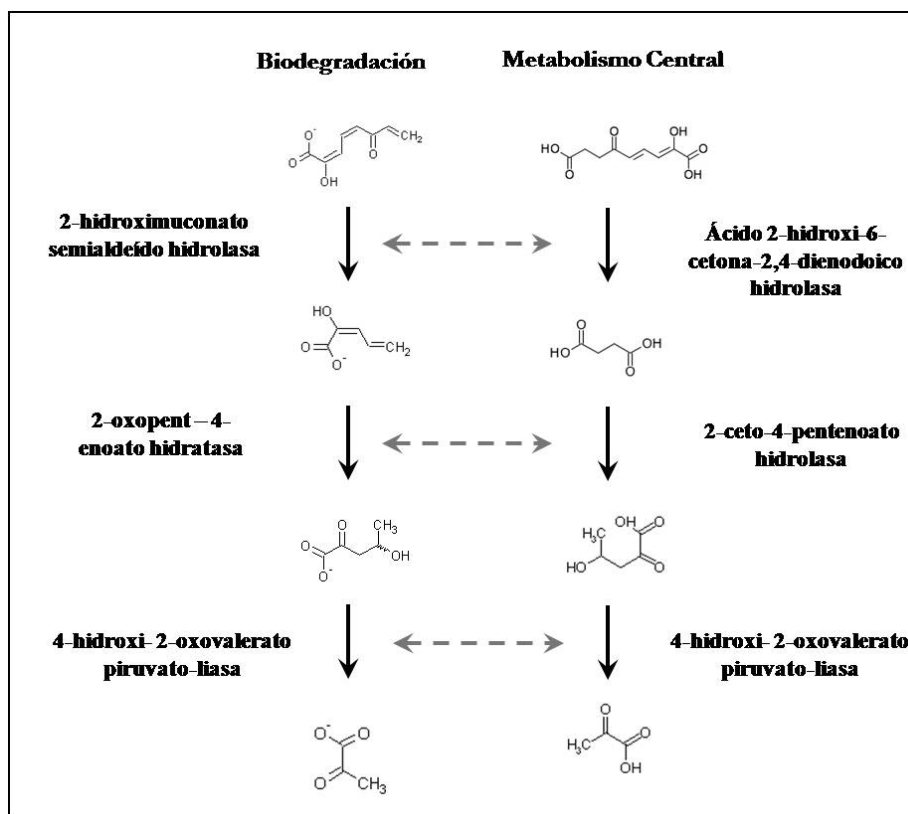


Figura 42. Ejemplo de bloques de reacciones consecutivas en biodegradación y el Metabolismo Central.

Para saber si la aparición de bloques relacionados evolutivamente con el Metabolismo Central es relevante o simplemente una consecuencia de la estructura de la red, he repetido mil veces el experimento anterior utilizando redes aleatorias análogas a la de biodegradación. En este caso se trata de redes donde se mantienen las conexiones de la red de biodegradación, el número de reacciones con relación evolutiva en el Metabolismo Central, el número de reacciones sin relación evolutiva en el Metabolismo Central y el número de reacciones con relación evolutiva mixta, pero la asignación del tipo relación a cada reacción se hace de manera aleatoria. La distribución de estas simulaciones muestra

que en la red de biodegradación el reclutamiento en serie de pares de reacciones con relación evolutiva en el Metabolismo Central es mucho mayor (con un z-score de 49,78) que lo que ocurre de manera aleatoria.

Tipo de Ensamblaje	Media	Desviación Típica
CM y en bloque	0.196	0.43
CM y no en bloque	6.598	2.89

Los resultados del estudio muestran que casi la totalidad de los pares de reacciones en biodegradación que siendo consecutivos mantienen una relación evolutiva cercana con reacciones del Metabolismo Central, han sido reclutadas en bloque (las reacciones del Metabolismo Central también son consecutiva) y aparecen en las parte final de la red de biodegradación.

4.6.5 Teorías de Evolución.

Los resultados anteriores parecen indicar que en muchos casos no existe una relación evolutiva cercana entre la red de biodegradación y la del Metabolismo Central. Como el 77% de las proteínas de biodegradación tienen al menos un homólogo cercano dentro de la propia red, se puede pensar que la formación y evolución de ésta surge desde su interior.

Existen varias teorías que han explicado la evolución de las redes metabólicas, siendo las principales el modelo de retro-evolución, el de duplicación de rutas y el de mosaico, detalladas en la introducción (Horowitz, 1945; Horowitz, 1965; Ycas, 1974; Jensen, 1976; Huynen & Snel, 2000; Rison & Thornton, 2002). Todas ellas están basadas en el estudio de pares de reacciones relacionadas por similitud de secuencia (a las que llamaré “RRSS”) y en la posición que ocupan en la red. Por tanto para estudiar estos mecanismos internos en la red de biodegradación y cuantificar su aportación en la expansión de la red, he identificado los casos de RRSS en la red de biodegradación, y para cada una de ellas calculado la distancia metabólica que separa a las reacciones.

Como ya he mencionado anteriormente, al ser los complejos que realizan una transformación homólogos entre sí, para analizar la homología de una reacción he asociado a cada una el complejo con mayor número de subunidades (el más completo), y de esta forma dos

reacciones serán RRSS si las proteínas de ambos complejos son homologas (para más detalle ver métodos).

La distancia entre dos reacciones viene dada por el número de pasos metabólicos que las separan. He fijado el umbral para decidir si dos reacciones son cercanas o lejanas en cinco pasos metabólicos basándome en que la mitad de las rutas de biodegradación tienen una longitud menor de cinco pasos siendo además estas rutas las que tienen más del 50% de sus reacciones con un metaenzima asociado. Para saber si el incremento de este umbral afectaría a los resultados, con un umbral igual a diez pasos metabólicos solo tendría cuatro casos más de reacciones de retro-evolución, que a continuación explicaré, no afectando en la estadística.

He considerado que un par de RRSS han surgido a partir de un evento de retro-evolución cuando este par de reacciones se encuentran cerca, están a menos de cinco pasos metabólicos (Figura 43).

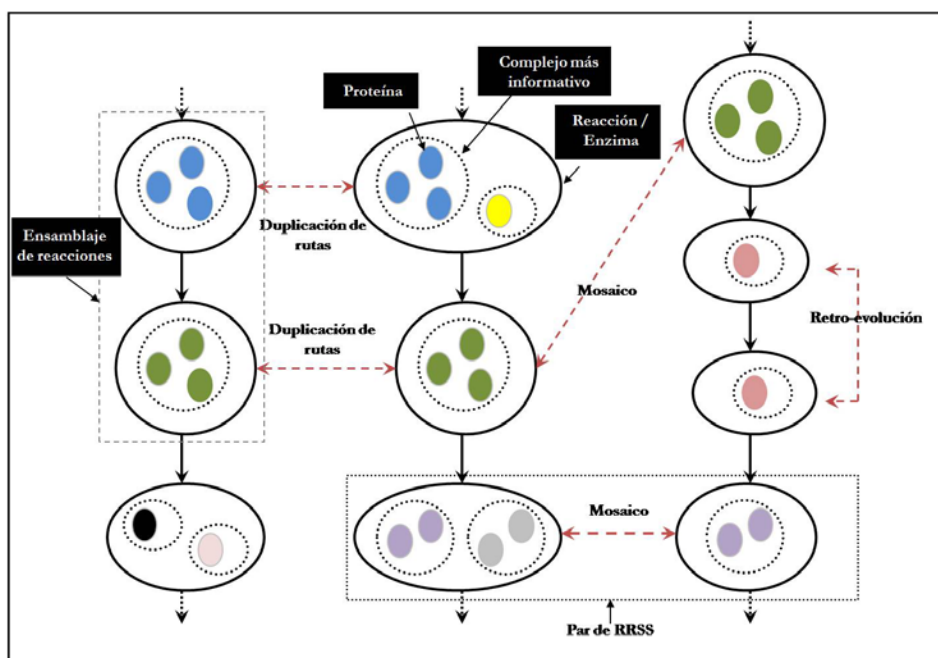


Figura 43. Conceptos y tipos de eventos evolutivos.

Si este par de RRSS son lejanas (hay una distancia mayor de cinco pasos entre ellas) y alguna de las reacciones consecutivas (entendidas como cualquier nodo (reacción) anterior o posterior en la red de biodegradación) en cada una de las reacciones del par forman también

otro par de RRSS, se ha considerado que el primero de los pares RRSS ha surgido a partir de un evento de duplicación de rutas. Por el contrario, si no se da ninguno de los casos anteriores, es decir, es un par de RRSS lejanas y sin otro par de RRSS consecutivo, entonces el primero de los pares de RRSS ha surgido a partir de un evento de “mosaico”.

En el caso de la red del Metabolismo Central se han encontrado ejemplos tanto de retro-evolución, por ejemplo la ruta de la glicolisis y del mandelato (Petsko *et al*, 1993), de duplicación de rutas, por ejemplo la biosíntesis de la histidina y el triptófano (Gerlt & Babbitt, 2001; Teichmann *et al*, 2001b; Teichmann *et al*, 2001a), como de mosaico, por ejemplo el ciclo de la urea (Copley, 2000; Teichmann *et al*, 2001a; Teichmann *et al*, 2001b). Sin embargo este último, el modelo de mosaico, parece ser el principal responsable en la evolución de nuevas rutas en la red del Metabolismo Central, seguido por el modelo de duplicación de pares y en raras ocasiones el modelo de retro-evolución (Teichmann *et al*, 2001a; Teichmann *et al*, 2001b).

He encontrado que el 70,5% de los pares de RRSS dentro de la red de biodegradación son lejanos y no tienen a su alrededor otro par de RRSS, siendo por lo tanto eventos de “mosaico”. Otro 27% de los pares de RRSS aún siendo lejanos tienen seguido otro par de RRSS, y que por lo tanto son eventos de “duplicación de rutas”. Por último, sólo un 2,5% de los pares de RRSS son cercanos y por lo tanto representan eventos de “retro-evolución” (Figura 44).

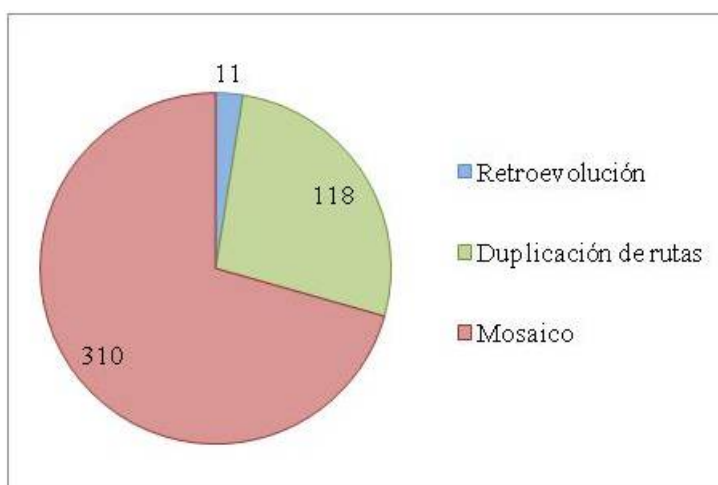


Figura 44. Representación de las teorías evolutivas en la red de biodegradación.

Para estudiar la distribución de cada uno de estos eventos a lo largo de la red de biodegradación, he definido la distancia del par de RRSS como la menor de las distancias de las dos reacciones que forman el par. Los resultados muestran que el 70% de los pares de RRSS metabólicamente cercanos, es decir eventos de retro-evolución, se encuentran a uno o dos pasos del Metabolismo Central. Un 70% de los eventos de mosaico y un 80% de los eventos de duplicación de rutas, se encuentran a distancias más alejadas, a tres o más pasos metabólicos del Metabolismo Central (Figura 45). Aunque el número de casos totales de retro-evolución es pequeño, he decidido analizar su distribución en la red para conocer la tendencia que podría seguir.

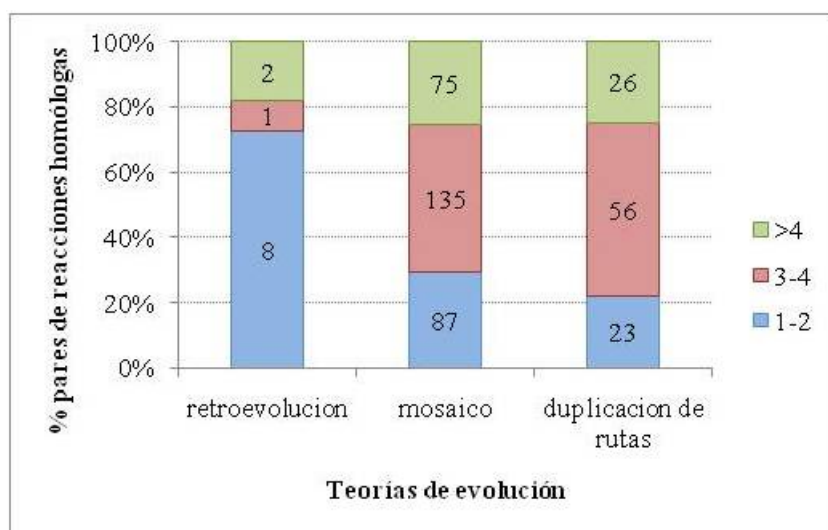


Figura 45. Distribución de las teorías evolutivas a lo largo de la red de biodegradación.

Por tanto, los resultados anteriores muestran por una parte la existencia de mecanismos internos en la red de biodegradación que participan en la formación y evolución de la propia red, siendo el más representativo el modelo en mosaico, seguido del modelo de duplicación de rutas y en raras ocasiones el modelo de retro-evolución. Este orden de frecuencia de aparición coincide con el observado anteriormente en el Metabolismo Central. Por otra parte, estos mecanismos de evolución no se distribuyen homogéneamente a lo largo de la red de biodegradación dándose más casos de retro-evolución en las proximidades del Metabolismo Central que en el resto de la red y por el contrario,

existiendo más eventos de mosaico y de duplicación de rutas en zonas alejadas del Metabolismo Central que en las proximidades de éste.

La “relación por similitud de secuencia” entre reacciones de biodegradación puede deberse a dos fenómenos: o que realmente las dos reacciones sean realizadas por la misma metaenzima (a lo cual he denominado dualidad funcional), o que el complejo de una de las reacciones se duplicara y más tarde debido a la introducción de mutaciones diverja para realizar la segunda reacción (a esto lo he denominado duplicación genética).

En la red de biodegradación, el 60% del total de los pares de RRSS se corresponden con eventos de duplicación genética. Si tenemos en cuenta el mecanismo que subyace detrás de cada uno de los pares de RRSS, obtengo que mientras que la duplicación genética es más habitual entre los pares con modelo de mosaico (se da en el 80% de estos) y modelo de retro-evolución (se da en el 60% de estos), la dualidad funcional es más abundante entre los pares con modelo de duplicación de rutas (se da en el 60% de estos) (Figura 46). Aunque el número de casos totales de retro-evolución es pequeño, he decidido analizar su RRSS para conocer la tendencia que podría seguir.

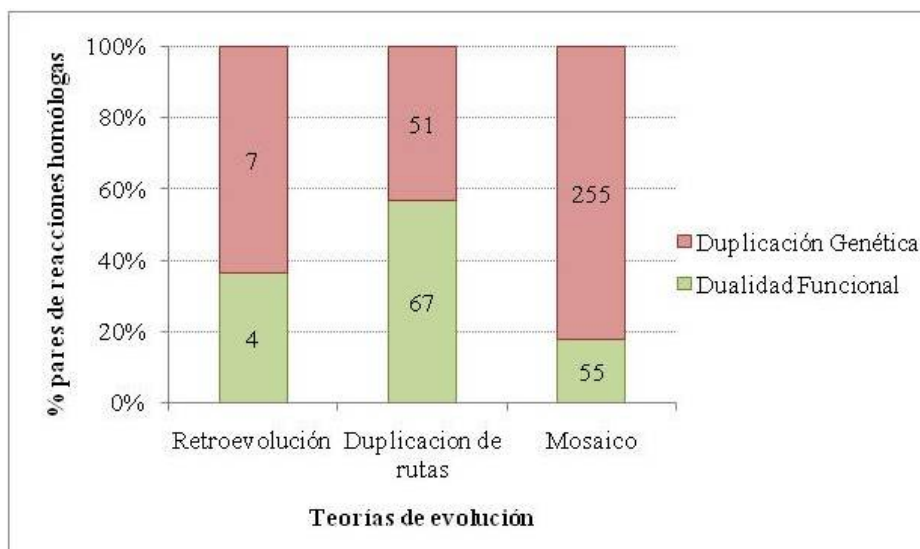


Figura 46. Representación de cada tipo de RRSS en las teorías evolutivas.

Casos en la red de biodegradación que son ejemplos de los resultados obtenidos son los siguientes. Los primeros pasos en la degradación del

tolueno y del 2,4-diclorotolueno pueden ser realizados por los mismos complejos de proteínas, tratándose por lo tanto de una duplicación de rutas mediante dualidad funcional. Por otro lado, los primeros pasos en la degradación del alcohol p-cúmico y del alcohol 4-sulfobencílico aunque también responde a un evento de duplicación de rutas, la duplicación en este caso es genética ya que los complejos que realizan las dos rutas no son los mismos pero sí homólogos. Por último, un evento de retro-evolución genética se da en las cercanías del Metabolismo Central donde el 2-bromomaleilacetato se transforma en maleilacetato y éste en 3-oxoadipato. Los complejos que realizan estas transformaciones aunque no son el mismo, sí que son homólogos (Figura 47).

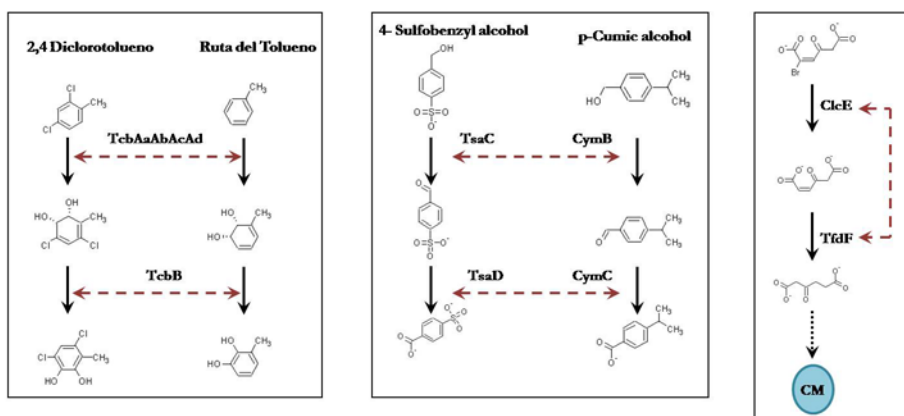


Figura 47. Ejemplos tipos de RRSS en la red de biodegradación. La gráfica de la izquierda representa un evento de duplicación de rutas mediante dualidad funcional. La figura central se corresponde con un evento de duplicación de rutas mediante duplicación genética. Por último el esquema de la derecha muestra un evento de retro-evolución por duplicación genética.

Basándome en los resultados obtenidos parece existir una tendencia a que en cada uno de los mecanismos que han contribuido en la evolución de la red de biodegradación un tipo de duplicación sea más abundante, siendo en el modelo de mosaico la duplicación genética y en los modelos de retro-evolución y duplicación de rutas la duplicación funcional.

4.7 Aplicación de la biología sintética: el detoxificador de arsénico.

La biología sintética al igual que la biología de sistemas, cuya utilización me ha permitido el estudio de la topología, organización y evolución de la red de biodegradación, abren nuevas perspectivas en el estudio de los sistemas biológicos. Si a esto se le suma mi predisposición a aplicar los conocimientos de electrónica, dada mi formación como ingeniero de telecomunicación, en el campo de la biodegradación, eje central de mi trabajo de doctorado, la utilización de la biología sintética en el desarrollo de una aplicación medioambiental es un paso lógico en la elaboración de mi tesis. Por ello, en esta segunda parte de mi tesis he desarrollado un sistema de detoxificación de arsénico.

El objetivo final es el diseño y construcción de un microorganismo capaz de ser un biorremediador frente a contaminaciones de arsénico en el medio. Para ello es necesario controlar el comportamiento de la bacteria en función del nivel de arsénico en el medio, capacitarla con la maquinaria necesaria para realizar la biorremediación y debido a la toxicidad del arsénico dotarla de un mecanismo de supervivencia.

Los mecanismos de defensa frente al arsénico son un sistema de bombeo, con el cual la bacteria expulsa de nuevo al medio el arsénico que ha absorbido, y un sistema de metilación del arsénico con el que lo metila convirtiéndolo en volátil y enviándolo al aire. Mientras el primero de los sistemas (él de bombeo) actúa como un mecanismo de supervivencia (al tomar la bacteria el arsénico del medio, éste lo devuelve de vuelta al mismo), el mecanismo de metilación es un sistema de biorremediación desde el punto de vista que elimina la cantidad de arsénico presente en el medio, ya que lo transforma en una especie volátil y menos tóxica que envía al aire.

La concepción del circuito de este proyecto se realizó en colaboración con Antonio David Páez del grupo de Víctor de Lorenzo, Centro Nacional de Biotecnología (CNB-CSIC) y bajo la supervisión de Ron Weiss, de la Universidad de Princeton. Una vez decidido el diseño mi labor continuaba en la realización del modelado y simulación del sistema, mientras que Antonio David Páez debería llevar a cabo la implementación experimental del circuito. Se decidió utilizar como

mecanismo de supervivencia el sistema de bombeo codificado en el operón “ars” de *Pseudomonas putida* (Canovas *et al*, 2003). El sistema de biorremediación elegido se corresponde con la metiltransferasa Cyt19 codificada en el gen *as3mt* presente en el hígado del ratón (Lin *et al*, 2002; Walton *et al*, 2003) y la cual se ha visto que es funcional en *Escherichia coli* (David Cánovas, comunicación personal).

La metiltransferasa Cyt19 es capaz de transformar el arsénico inorgánico en dimetil-arsenato, un compuesto menos tóxico (Walton *et al*, 2003). Por su parte, el operón “ars” de *Pseudomonas putida* está formado por los genes *arsR*, *arsB*, *arsC*, *arsH* y se activa por la acción del promotor Pars (Canovas *et al*, 2003). El modelo molecular y de regulación de este operón es el siguiente. La proteína ArsR reprime a su promotor impidiendo la expresión del operon “ars”. Cuando se introduce en el medio arsénico, este se une a la proteína ArsR inhibiendo la represión y expresándose por tanto el operon “ars” (es decir, las proteínas ArsR, ArsB, ArsC y ArsH). Si el arsénico se presenta en forma de arsenitos (As III) la proteína ArsB es capaz de bombearla directamente al exterior. Por el contrario si el arsénico está en forma de arsenato (As V) es necesario que la proteína ArsC lo transforme en arsenito (As III) y de esta forma poder ser bombeada al exterior por ArsB. La función de la proteína ArsH es desconocida hasta el momento (Neyt *et al*, 1997) (Figura 48).

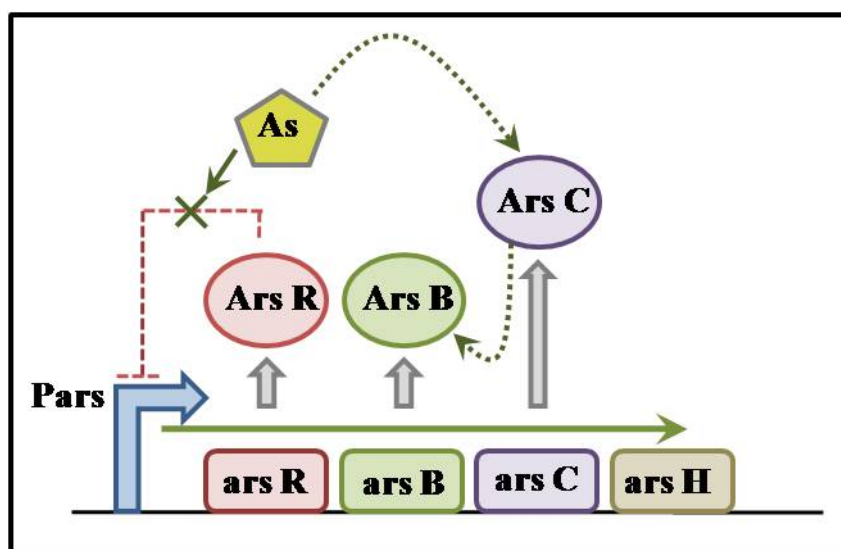


Figura 48. Descripción del operón Ars. Los cuadrados representan genes, los óvalos proteínas, la flecha azul el promotor Pars, el pentágono el compuesto de arsénico, las flechas grises la expresión de los genes en proteínas, la flecha verde la dirección en la

que transcribe el promotor, las flechas verdes discontinuas la sucesión de eventos tras la aparición del arsénico. La flecha roja discontinua indica la represión de la proteína ArsR sobre su promotor Pars. Con la presencia del arsénico, esta represión se elimina.

4.7.1 Diseño del detoxificación.

Dado que los mecanismos anteriores cumplen roles diferentes, el bombeo supervivencia y la metilación biorremediación, la activación de cada uno de ellos debería de ser diferente y estar condicionada a la cantidad de arsénico existente. Mientras que el mecanismo de metilación sería interesante que estuviera activo con la simple presencia de arsénico, independientemente de su cantidad, el mecanismo de bombeo solo sería necesario cuando el nivel de arsénico comenzara a ser tóxico para la bacteria. Este comportamiento podría conseguirse utilizando a la entrada de los mecanismos un comparador que decidiera que mecanismos tener activo, sólo uno o los dos simultáneamente, tras comparar su señal de entrada con un cierto umbral que tendría fijado.

Por lo tanto la sucesión de eventos que se darían en el sistema sería la siguiente. Los arsenatos (As V) del medio son absorbidos por el organismo a través de los canales de fosfatos y los arsenitos (As III) a través de porinas de glicerol. La cantidad de arsenatos y arsenitos que se introducen dentro del organismo hacen disminuir la cantidad total de estos compuestos en el medio. Los arsenatos en el organismo se transforman en arsenitos mediante la acción de la proteína Ars C. La cantidad total de arsenito en el organismo (la que ha tomado del medio y la que se ha transformado a partir de arsenato) se compara con un cierto umbral y se produce una respuesta en función del resultado de la comparación. A partir de un nivel muy pequeño de arsenito se activa el mecanismo de metilación convirtiendo el arsenito en dimetil-arsenato y expulsándolo al aire. Si el nivel de arsenito es mayor, simultáneamente al sistema de metilación se activa el de bombeo, devolviendo parte del arsenito al medio (Figura 49).

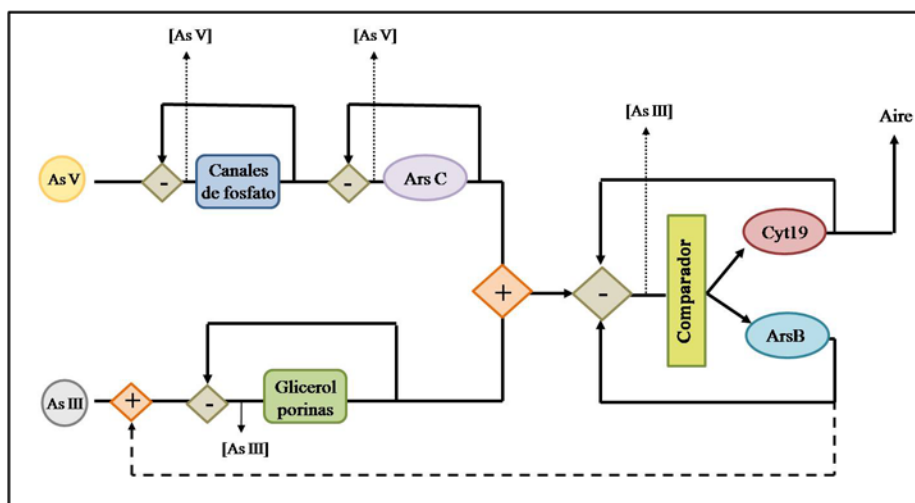


Figura 49. La representación electrónica de cada uno de las partes del circuito. Los sumadores, signo "+", restadores, signo "-", adquisición por canal de fosfato o porinas de glicerol, el comparador y los mecanismos de bombeo, ArsB y metilación, Cyt19. Podría llevarse a cabo utilizando amplificadores operacionales en sus diferentes configuraciones (Figura 50).

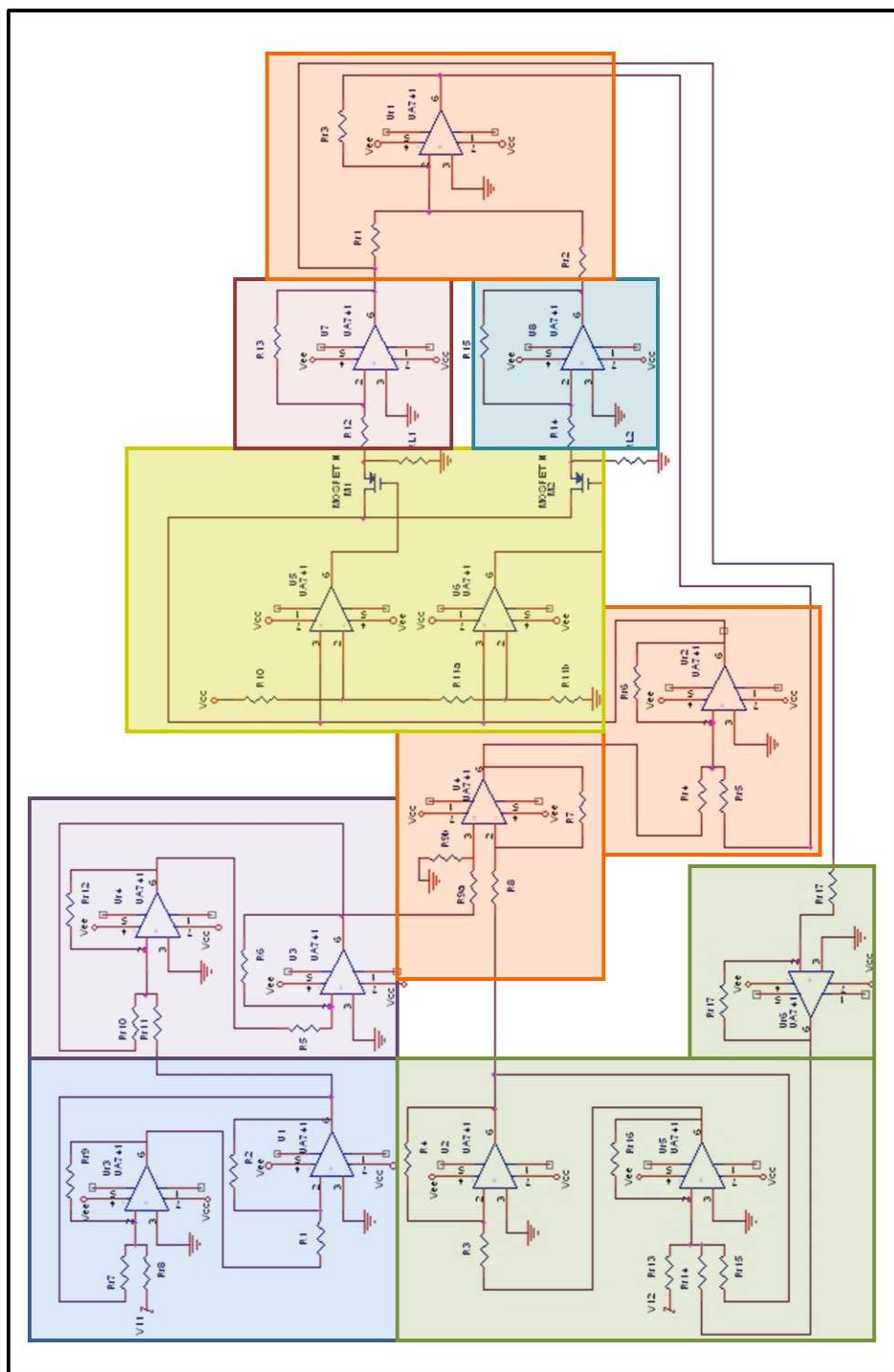


Figura 50. Esta figura muestra el diseño electrónico que se comportaría de la manera mostrada en la Figura 49. Los colores en esta figura se corresponden con

cada uno de los elementos de la figura 49. Los triángulos son amplificadores operacionales, los fragmentos de línea azul en zig-zag representan resistencias, Cada amplificador operacional debe estar conectado a unas tensiones de referencia (+Vcc y -Vcc) para su correcto funcionamiento. El triángulo formado por líneas discontinuas representa masa o tierra.

Dada la complejidad del sistema como primera aproximación parto de la cantidad de arsénico en el interior del organismo, sin diferenciar entre arsenato y arsenito.

El diseño biológico del comparador se ha basado en el trabajo realizado por el grupo de Ron Weiss de la Universidad de Princeton, en el cuál utilizando los pares represor-promotor cI-P λ y lacI-Plac, y modificando un codón de *lacI* que reduce la eficiencia de represión de la proteína LacI sobre el promotor Plac, consiguen dos niveles de activación distintos para cada par represor-promotor (Jacob & Monod, 1961; Ptashne, 2004; Basu *et al*, 2005). Por otra parte, el sensor del detoxificador de arsénico que dispara el funcionamiento del circuito en presencia del arsénico se consigue utilizando la represión de la proteína ArsR sobre el promotor Pars y la inhibición de esta represión en presencia de arsénico.

Por tanto el funcionamiento ideal del sistema sería el siguiente. Cuando la cantidad de Arsénico es suficiente para desbloquear la inhibición de ArsR sobre Pars, el mecanismo de metilación se activa. Sin embargo solamente cuando el nivel de Arsénico aumente y la cantidad de proteína LacIM (proteína LacI modificada) sea la suficiente como para reprimir al promotor Plac (ya que al ser menos eficiente en la represión necesita mayor cantidad), reprimirá la represión de la proteína CI sobre P λ permitiendo que se active el bombeo (Figura 51).

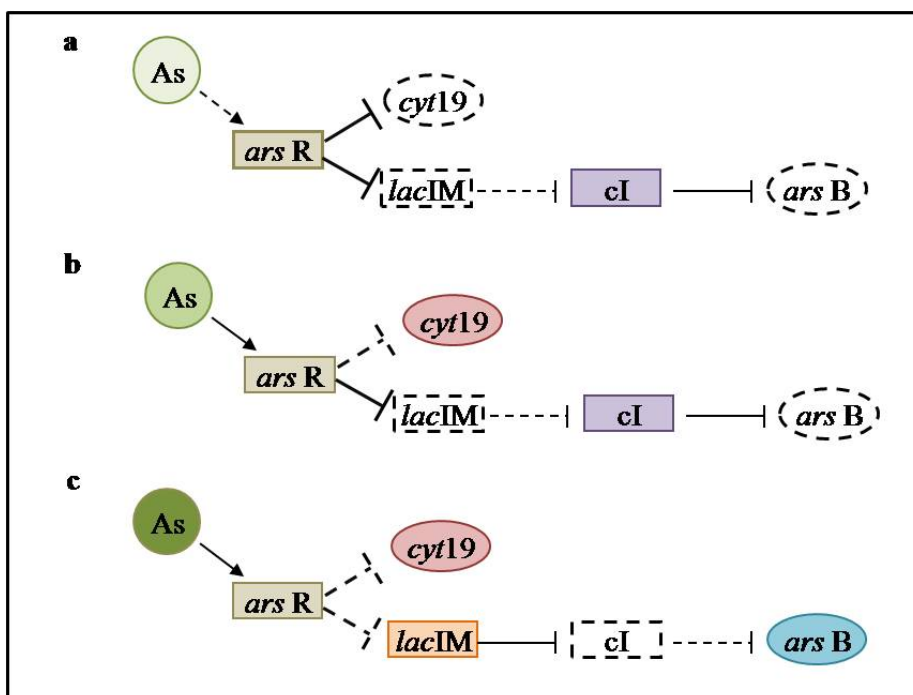


Figura 51. Esquema del funcionamiento del detoxificador. a) Cuando no haya arsénico, el represor ArsR inhibe la expresión de la metilasa directamente y de la bomba de manera indirecta. b) cuando la concentración de arsénico aumente, la represión se reducirá, expresándose la metilasa. Sin embargo la represión todavía es la suficiente para inhibir el funcionamiento de la bomba. c) cuando la concentración de arsénico sea tóxica para el organismo, la represión de ArsR es tan débil que tanto la metilasa como la bomba se expresan. La activación o represión de la bomba se obtiene usando una cadena de represores y modificando la proteína LacI para obtener LacIM cuya eficiencia de represión sobre Plac es menor.

La construcción biológica que se comporta de la manera descrita, consta de un promotor constitutivo (*PlacIq*) que hace que siempre exista proteína ArsR en cantidades suficientes como para bloquear al promotor *Pars*. A su vez el promotor *Pars* es el que expresa la metilasa y se encarga de desencadenar una cadena de dos represores (*LacIM* --| *CI* --| *ArsB*) que se comportan como un activador. Por tanto *LacIM* reprime al promotor *Plac* que es el que expresa al represor *CI*, que a su vez reprime al promotor *Pλ*, encargado de expresar el mecanismo de bombeo. En el diseño también se han incluido dos proteínas que al expresarse emiten fluorescencia, la proteína GFP con una respuesta verde y la XFP con una roja, para que al ser implementado el mecanismo experimentalmente pueda supervisarse la expresión de cada uno de los mecanismos (Figura 52).

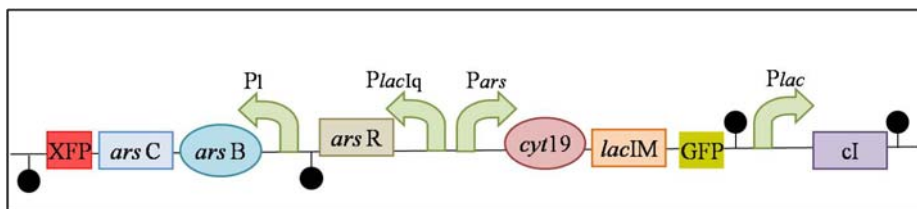


Figura 52. Diagrama del circuito biológico del detoxificador de arsénico. En él se muestran el mecanismo de supervivencia (*arsB*) y biorremediación (*cyt19*) y la regulación que controla sus funcionamientos. Las proteínas fluorescentes GFP y XFP se han incluido en el modelo para que al ser implementado sea fácil la detección del mecanismo activo en cada instante. Los círculos negros representan los terminadores de la transcripción de cada uno de los promotores.

4.7.2 Modelado del detoxificador.

El modelado de este sistema lo he realizado utilizando la herramienta de Matlab “SimBiology”, la obtención de los parámetros y los detalles de esta se describen en el apartado Materiales y Métodos. En el modelado, se han establecido las relaciones entre cada uno de los elementos del circuito biológico, incluyendo los eventos de transcripción del ADN en ARN mensajero (ARNm), de traducción del ARNm en proteínas, de dimerización de las proteínas, de unión de los represores a sus promotores, de degradación de las proteínas, y el orden en que se dan estos eventos. Se pueden encontrar más detalles sobre los valores por defecto asignados a cada uno de estos eventos, y las cantidades iniciales de los elementos del sistema en la sección de Materiales y Métodos. A continuación se muestra el sistema que se forma utilizando los elementos y relaciones descritas (Figura 53).

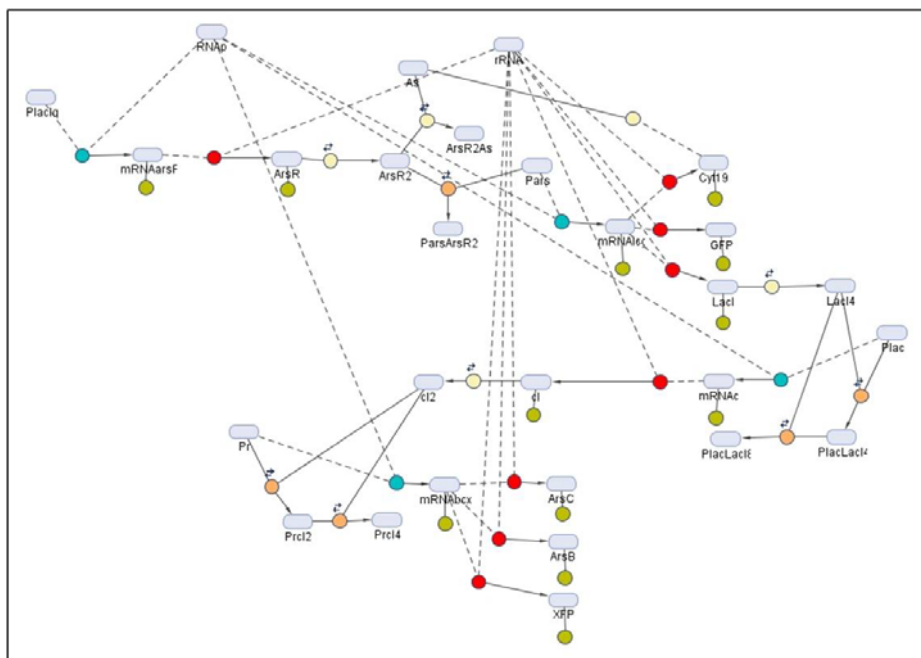


Figura 53. Representación gráfica de las reacciones del circuito global. En la figura las especies se representan con cuadrados azules, las transformaciones con círculos, los reactantes con líneas discontinuas y los productos con flechas continuas. Cuando una transformación es reversible se indica con dos flechas sobre el círculo.

Al ser el sistema tan complejo con un gran número de parámetros, lo he dividido en tres módulos donde se pueden estudiar mejor como afectan los parámetros del sistema en la salida que idealmente quiero conseguir. El primer módulo abarca el circuito que actúa como sensor de arsénico y disparador tanto del mecanismo de metilación (una de las salidas de este módulo) como de la represión de la cadena de represores que inhibe a la bomba. El segundo módulo incluye la represión de la proteína LacIM sobre el promotor Plac, y el tercer módulo refleja represión de la proteína CI sobre el promotor P_{λ} y por lo tanto sobre el mecanismo de bombeo (Figura 54).

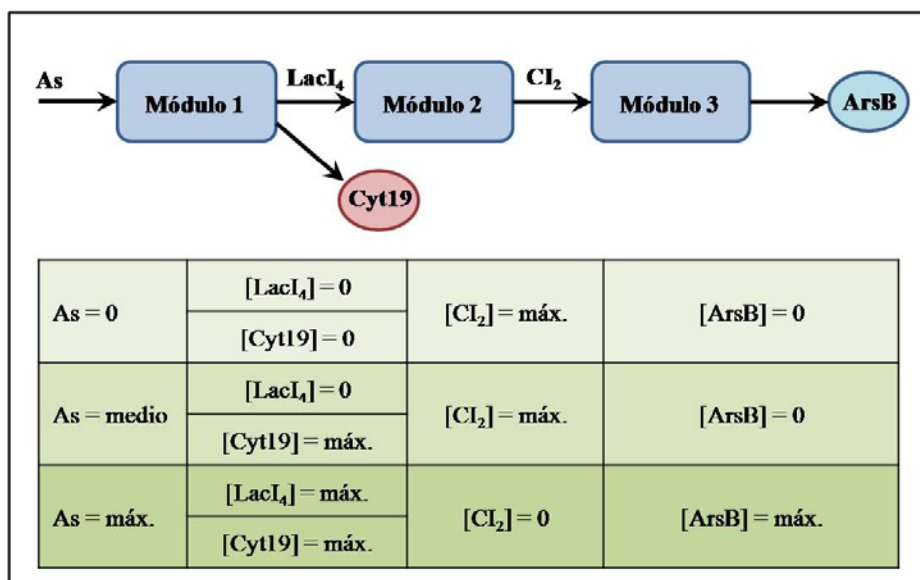


Figura 54. Modularización del sistema. El primer módulo tiene como entrada la concentración de arsénico y da como salida la cantidad de metilasa en el sistema y la cantidad del tetramero de proteínas LacI. Para que LacI reprima al promotor Plac es necesario la formación de dos tetrámeros que se unen al promotor inhibiendo su expresión. El segundo módulo tiene como entrada la cantidad de tetrámeros LacI (LacI₄) y da como salida el número de dímeros de la proteína CI (CI₂). Para reprimir al promotor P_λ se necesitan dos dímeros de CI que bloqueen su expresión. El tercer módulo tiene como entrada la cantidad de dímeros CI (CI₂) y como salida la expresión de la bomba de arsénico. En La tabla se muestran los niveles deseados de cada una de las entradas y salidas de los módulos para que el comportamiento del sistema sea el deseado.

El objetivo es conseguir que cuando no existe arsénico ni la metilación ni el bombeo estén activos, que cuando el nivel de arsénico no sea tóxico para la bacteria solamente actúe la metilación, pero cuando el nivel de arsénico ponga en riesgo la supervivencia del organismo, se active además de la metilación el mecanismo de bombeo.

4.7.3 Simulación del detoxificador.

Para fijar los valores de arsénico que activan cada rango de funcionamiento sería necesario medirlo experimentalmente. En la ausencia de estos datos voy a realizar una aproximación estudiando el comportamiento del sistema si se varía la cantidad de arsénico. Del mismo modo ante la imposibilidad de obtener valores precisos para los parámetros de cada una de las reacciones del modelo, se ha optado por

utilizar unos por defecto. Las dos premisas anteriores no suponen ningún problema ya que con la simulación del sistema no trato de cuantificarlo sino de conocer si cada uno de los módulos funciona de la manera deseada e identificar que parámetros son claves (tasas de transcripción, de traducción, de degradación, ...) en la respuesta del sistema. La información obtenida en la simulación es por tanto de utilidad para que al construir el circuito de manera experimental se tenga una idea de que elementos y como han de modificarse. Sin embargo a medida que se tenga más datos experimentales, se podrá ir ajustando los parámetros para estudiar el sistema con más detalle.

Las modificaciones que pueden realizarse experimentalmente son principalmente alterar la tasa de degradación de una proteína, la fuerza con la que un promotor expresa sus genes, la fuerza del sitio de unión al ribosoma (el cual afecta a la eficiencia de traducción del mRNA en proteína) y la afinidad de unión entre un represor y el promotor al que reprime. Por ello, solo propondré modificaciones en los parámetros implicados en estos procesos.

4.7.3.1 Primer módulo: Sensor de arsénico y metilación.

La figura 55 muestra un esquema de este primer módulo. El promotor constitutivo $PlacI_q$ se une a la RNA polimerasa (RNAP) y sintetizan el $mRNA_{ArsR}$ el cuál puede degradarse y desaparecer o ser decodificado por el ribosoma (rRNA) para dar lugar a la proteína ArsR. Esta proteína puede degradarse y desaparecer o agruparse en un dímero ArsR2. Estos dímeros pueden volverse a desagrupar en proteínas individuales (ArsR), unirse a una molécula de arsénico (As) para formar el compuesto ArsR2As o unirse al Promotor Pars y reprimir la expresión de sus genes (ParsArsR2). El compuesto ArsR2As a su vez puede desagruparse dando lugar a arsénico y ArsR2. El promotor Pars (al no estar inhibido) se une a la RNAP para producir el $mRNA_{lsg}$ que será decodificado por el ribosoma en las proteínas Cyt19, GFP, y LacI. La especie “ParsArsR2” puede desligarse dando lugar a Pars y ArsR2. Las proteínas Cyt19, GFP y LacI se pueden degradar y desaparecer. A su vez LacI forma tetrámeros que serán los que repriman la expresión del promotor Plac en el módulo segundo. Estos tetrámeros pueden desagruparse dando lugar a proteínas LacI. La unión del dímero ArsR2 con el arsénico o el con el promotor Pars es lo que hace que el comportamiento del sistema cambie entre ser un represor o no.

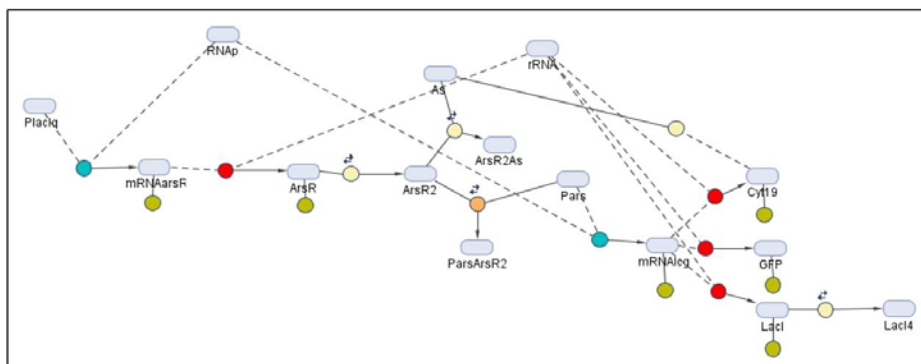


Figura 55. Representación de las reacciones del primer módulo. Los círculos verdes representan las transformes donde puede alterarse la tasa de degradación de una proteína, los azules las modificaciones que se pueden realizar a la fuerza del promotor, los rojos la modificaciones de la fuerza del Sitio de Unión al Ribosoma y los naranjas afinidad de unión entre el represor y el promotor. Las flechas de color azul muestran los parámetros a los que es más sensible este módulo y cuyo incremento producen una reducción de la salida del módulo, mientras que las flechas verdes indican los parámetros a los que es más sensible este módulo y cuyo incremento producen un aumento de la salida del módulo.

El comportamiento ideal consiste en que cuando no hay arsénico no exista ni Cyt19 ni LacI4, y que a medida que el nivel de arsénico incrementa, aumente la cantidad de ambos, siendo el incremento de Cyt19 mayor que para LacI4. Utilizando los parámetros definidos por defecto para cada uno de los eventos (tasa de degradación, dimerización, traducción, etc.) y variando el nivel de arsénico el comportamiento del módulo se refleja en la figura 56.

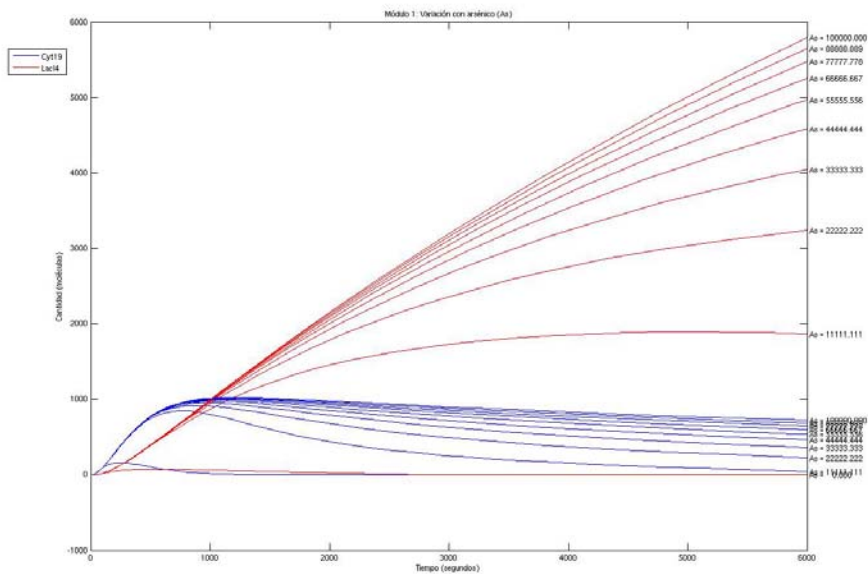


Figura 56. Función de transferencia del primer módulo con los parámetros por defecto. Respuesta de las salidas proteína Cyt19 (línea azul) y tetrámero LacI4 (línea rojo) ante la variación de la cantidad arsénico.

Como se puede observar, el sistema responde de la manera deseada, sin embargo la cantidad de LacI4 aumenta más rápido que Cyt19. Para solventar este problema he probado por una parte a aumentar la tasa de traducción de Cyt19 en un nivel de magnitud (de 0.033 a 0.33) (Figura 57) y por otra parte a disminuir un nivel de magnitud la tasa de degradación de la proteína Cyt19 (de 0.0050 a 0.00050) (Figura 58).

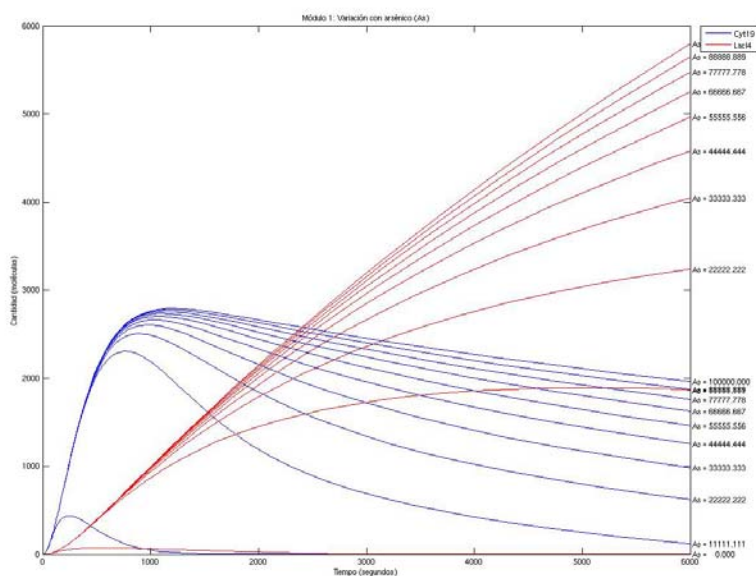


Figura 57. Función de transferencia del primer módulo tras aumentar la tasa de traducción de Cyt19. Respuesta de las salidas proteína Cyt19 (línea azul) y tetrámero LacI4 (línea rojo) ante la variación de la cantidad arsénico.

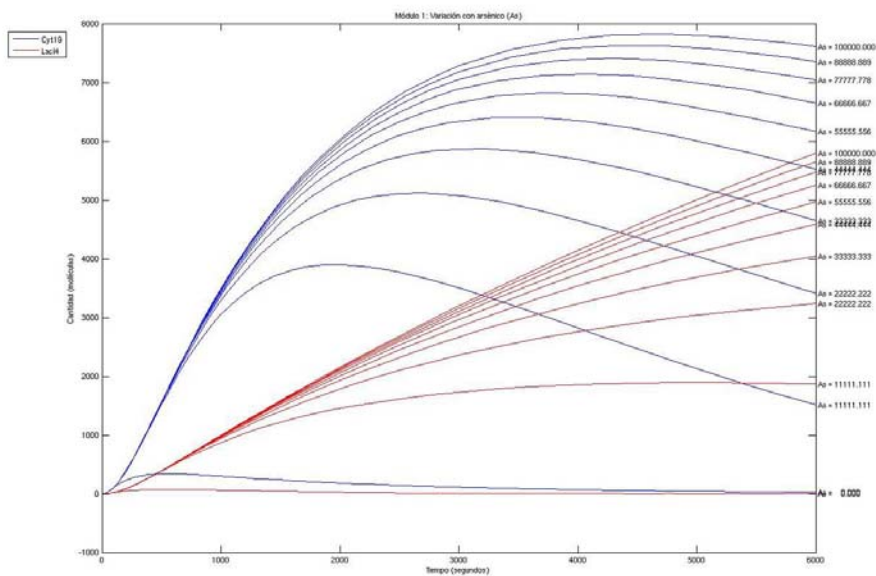


Figura 58. Función de transferencia del primer módulo tras disminuir la tasa de degradación de la proteína Cyt19. Respuesta de las salidas proteína Cyt19 (línea azul) y tetrámero LacI4 (línea rojo) ante la variación de la cantidad arsénico.

4.7.3.2 Segundo módulo: Represión de CI.

El segundo módulo representado en la figura 59 engloba desde el tetrámero de LacI (LacI4) hasta la formación del dímero de CI (CI2). El tetrámero LacI4 se une a dos sitios del promotor Plac para inhibir su expresión formando PlacLacI4 y PlacLacI8. PlacLacI4 puede desligarse y dar lugar a LacI4 y Plac. A su vez PlacLacI8 puede desligarse y producir LacI4 y PlacLacI4. El promotor Plac (al no estar inhibido) se une a la RNAP para dar lugar al mRNAci, el cuál puede o degradarse y desaparecer o ser decodificado por el rRNA en la proteína CI. La proteína CI puede degradarse y desaparecer o formar dímeros (CI2) que serán los encargados de reprimir la expresión del promotor Pr en el módulo tercero. Este dímero también puede desagruparse para dar lugar a proteínas CI.

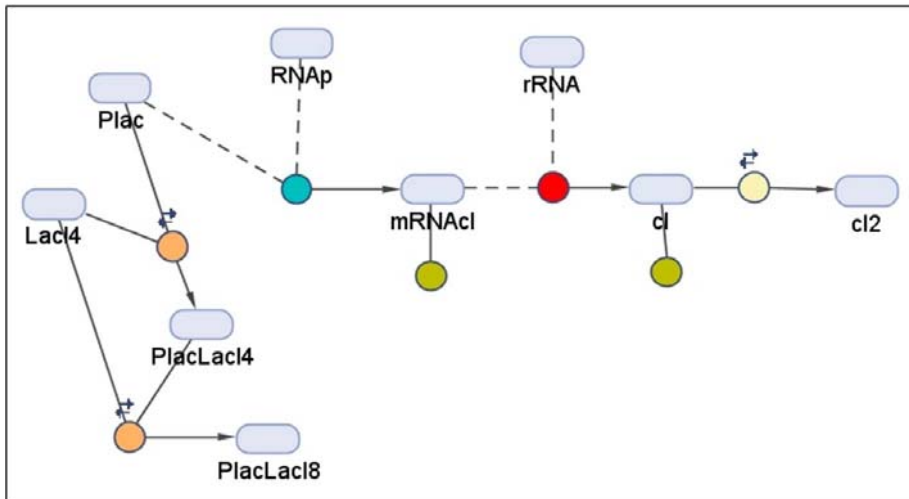


Figura 59. Representación de las reacciones del segundo módulo. Los colores de los círculos y las flechas son los mismos que los descritos en el esquema del primer módulo (Figura 55).

El comportamiento ideal de éste segundo módulo implica que cuando no haya LacI4, haya mucha cantidad del dímero CI2 pero que cuando haya LacI4, no haya CI2. Utilizando los parámetros por defecto y variando la cantidad de LacI4, la función de transferencia del segundo módulo se representa en la figura 60. En la gráfica se observa que

únicamente cuando la cantidad de LacI4 es 0, en la salida del sistema se tiene CI2. El funcionamiento del represor es correcto, por tanto no es necesario introducir ningún tipo de modificación en los parámetros.

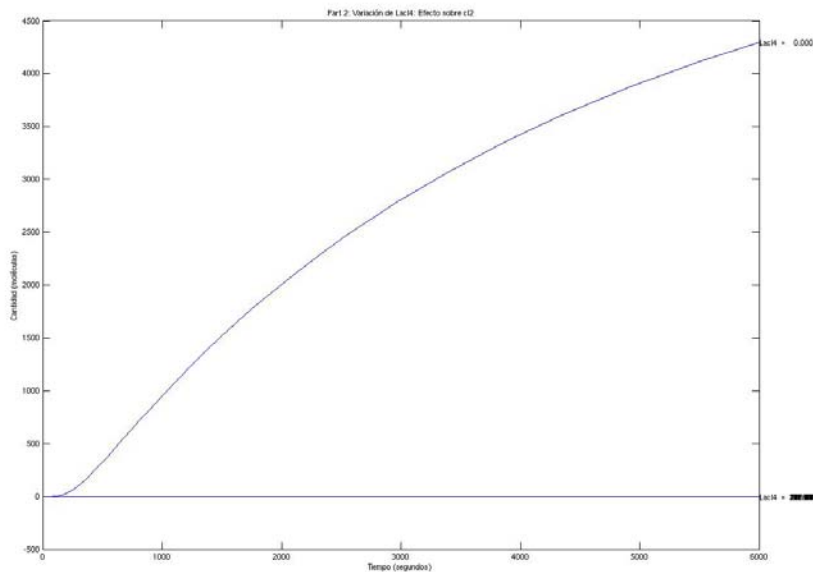


Figura 60. Función de transferencia del segundo módulo con los parámetros por defecto. Respuesta del dímero CI2 ante la variación de la cantidad de LacI4.

4.7.3.3 Tercer módulo: Activación de la bomba de arsénico.

Este último módulo representado en la figura 61 tiene como entrada la cantidad de dímero CI2 y como salida la cantidad de proteína ArsB (la bomba). El dímero CI2 se une por dos sitios al promotor Pr para reprimirlo produciendo PrCI2 y PrCI4. PrCI4 puede desligarse para dar lugar a CI2 y PrCI2. A su vez PrCI2 puede desligarse en Pr y CI2. El promotor Pr (al no estar inhibido) se une a la RNAP para dar lugar al mRNA_{bcx}, el cual puede o degradarse y desaparecer o ser decodificado por el rRNA en las proteínas ArsB, ArsC y XFP. Las proteínas ArsB, ArsC y XFP pueden degradarse y desaparecer.

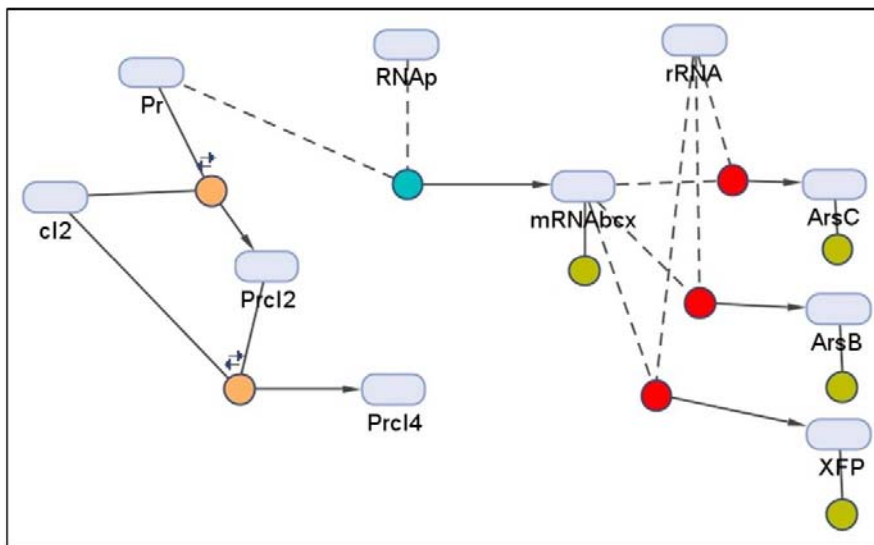


Figura 61. Representación de las reacciones del tercer módulo. Los colores de los círculos y las flechas son los mismos que los descritos en el esquema del primer módulo (Figura 55).

El comportamiento ideal de este módulo implica que cuando no haya **CI2**, exista mucha cantidad de **ArsB** y que cuando haya **CI2**, no haya nada de **ArsB**. En la figura 62 se muestra el resultado de la simulación del módulo con los valores por defecto de los parámetros y variando la cantidad del dímero **CI2**. En la gráfica se observa que sólo cuando la cantidad de **CI2** es 0, el sistema produce una salida (la cantidad de **ArsB** es diferente de 0). El funcionamiento del represor es correcto, por tanto no es necesario introducir ningún tipo de modificación en los parámetros.

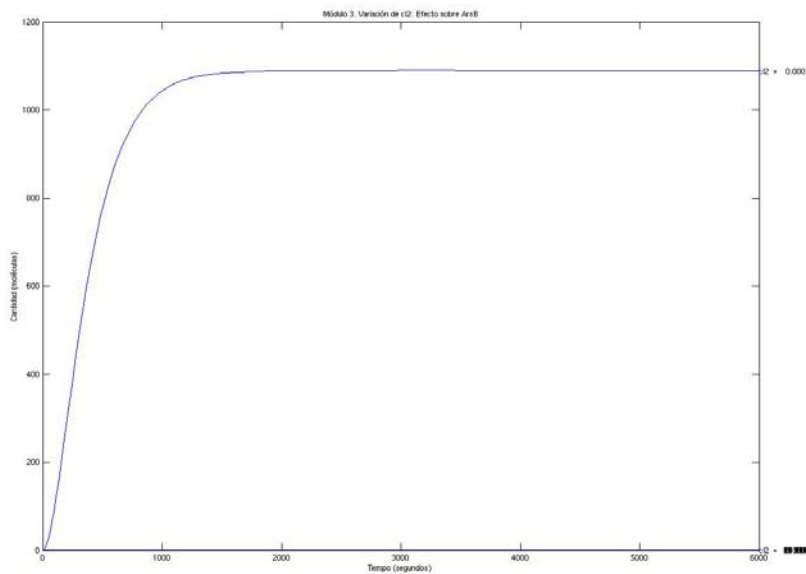


Figura 62. Función de transferencia del tercer módulo con los parámetros por defecto. Respuesta de la proteína ArsB ante la variación de la cantidad de CI2.

4.7.3.4 Funcionamiento global del sistema.

Una vez que se ha comprobado el funcionamiento de cada uno de los módulos, el siguiente paso es el estudio del comportamiento global del circuito. La figura 63 muestra como varían los mecanismos de metilación (Cyt19) y de bombeo (ArsB) en función de la variación de la cantidad de arsénico utilizando los parámetros por defecto excepto para la tasa de degradación de la proteína Cyt19, la cual se había disminuido en un grado de magnitud.

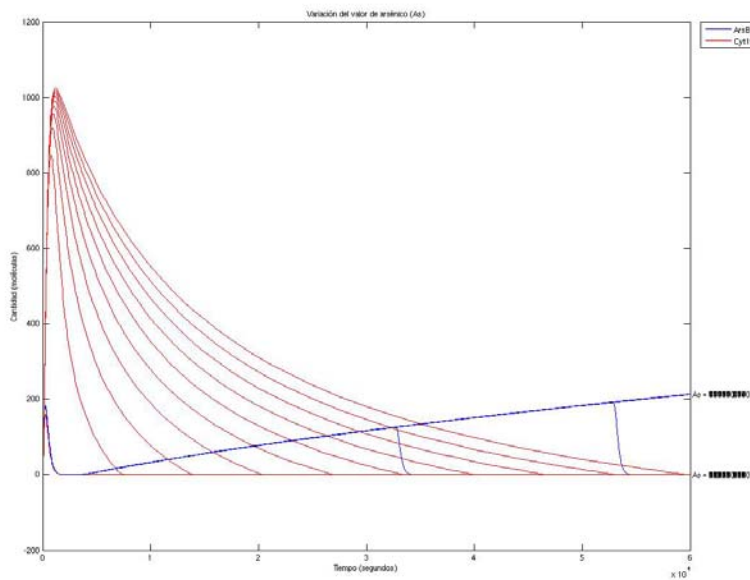


Figura 63. Función de transferencia del sistema global tras disminuir la tasa de degradación de la proteína Cyt19. Respuesta de las salidas proteína Cyt19 (línea roja) y proteína ArsB (línea azul) ante la variación de la cantidad arsénico.

En la gráfica la cantidad de Cyt19 es muy superior a la de ArsB. Para solventar este problema he disminuido la tasa de degradación y aumentado la de traducción de ArsB, ambas en un grado de magnitud. El resultado se muestra en la Figura 64.

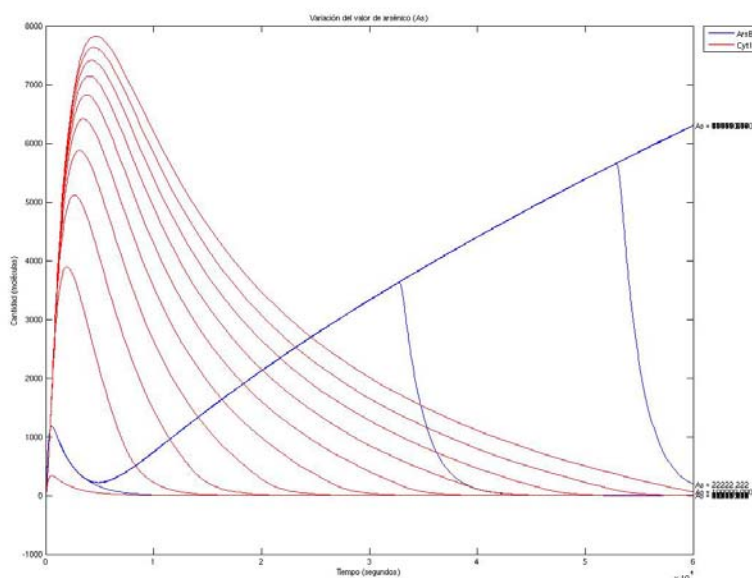


Figura 64. Función de transferencia del sistema global tras disminuir la tasa de degradación de la proteína Cyt19, disminuir la tasa de degradación de la proteína ArsB y aumentar la tasa de traducción de ArsB. Respuesta de las salidas proteína Cyt19 (línea roja) y proteína ArsB (línea azul) ante la variación de la cantidad arsénico.

Tanto en la figura 63 como en la figura 64 se observa que el circuito presenta un retraso importante a la hora de proporcionar la salida del último módulo, cantidad de ArsB. Para analizar este problema, en la Figura 65 se muestran la variación de las entradas y salidas de los módulos del sistema frente a la variación de arsénico, introducidas las modificaciones comentadas anteriormente (disminución de la tasa de degradación de Cyt19 y ArsB y aumento de la tasa de traducción de ArsB). En la gráfica se observa como para que se produzca ArsB es necesario que haya LacI4 para reprimir la producción de CI2 y que desaparezca la cantidad existente de CI2. El comportamiento deseado se correspondería con una activación de ArsB inmediata en el momento que la cantidad de arsénico fuera suficiente y cuya pendiente de crecimiento fuera prácticamente perpendicular, como se observa para la proteína Cyt19.

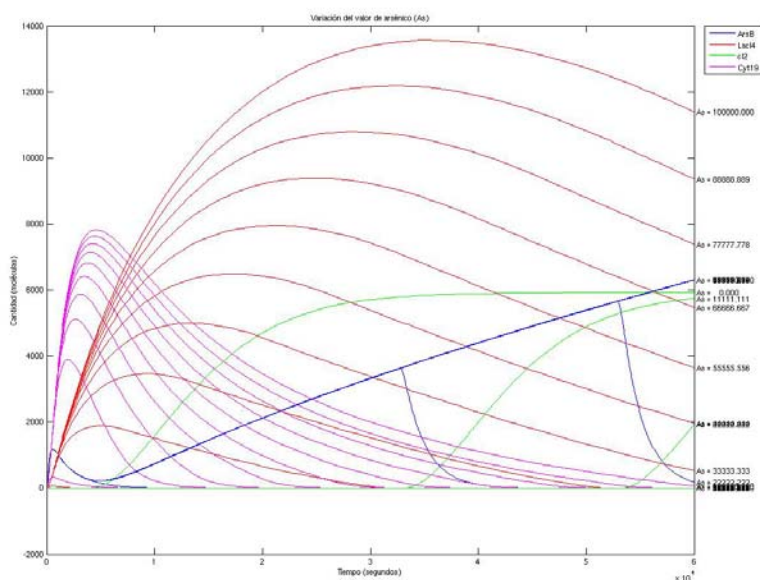


Figura 65. Análisis de los retrasos en la respuesta del sistema global. Variación de las entradas y salidas de los módulos del sistema frente a la variación de arsénico. En este sistema se ha disminuido la tasa de degradación de Cyt19 y ArsB y aumentado la tasa de traducción de ArsB.

4.7.3.5 Posibles configuraciones que mejorarían el sistema.

En el sistema la activación de la bomba se consigue mediante dos pares de represor-promotor en cadena. Esta configuración produce retrasos en la respuesta del sistema de bombeo ya que aunque el nivel de arsénico sea suficiente para producir la cantidad necesaria de la proteína LacI que reprime al represor CI que reprime a su vez a la bomba, es necesario esperar a que la cantidad existente de CI desaparezca. Para evitar este problema sería posible sustituir los dos pares represor-promotor en cadena por un activador-promotor, por ejemplo el formado por la proteína AraC y su promotor *Para*. El inconveniente del uso de este activador es que para que AraC funcione como un activador, es necesario que esté unido a Arabinosa(Englesberg *et al*, 1965; Greenblatt & Schleif, 1971), y por lo tanto se tendría que incorporar en el medio este compuesto, limitando la aplicabilidad del detoxificador (Figura 66).

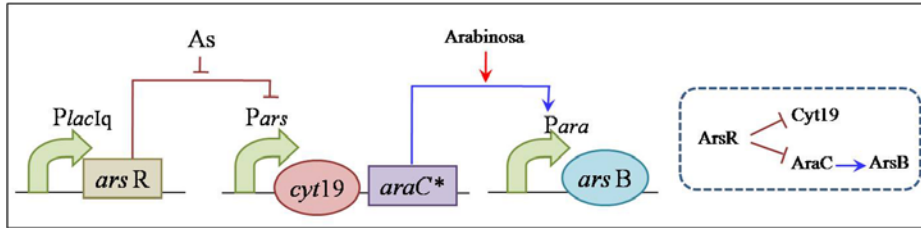


Figura 66. Configuración: mejora en la velocidad de respuesta del sistema. La construcción se diferencia del circuito original en utilizar un activador (*araC*) en lugar de dos represores en serie (*LacI* y *CI*). El asterisco de *araC* indica que el nivel de activación deseado del sistema de bombeo se obtiene modificando la eficiencia de activación de la proteína *AraC* sobre el promotor *Para*.

Por otra parte, la manera en que hemos conseguido fijar un nivel de activación de la bomba mayor que el de la metilación es mediante la mutación de la proteína *LacI*, haciéndola mucho menos eficiente en su represión sobre *Plac*. Otra manera de conseguir esta diferencia en la activación implicaría el uso del promotor *Pars* tanto para activar el funcionamiento de la metilación como para activar la bomba, pero en este último el promotor estaría modificado para ser más sensible a la presencia de *ArsR* y por tanto requerir mayor cantidad de moléculas de arsénico que bloqueen la acción de *ArsR* sobre el promotor. Con esta configuración es más sencillo ajustar el nivel de activación de cada mecanismo ya que son independientes a diferencia del circuito original (Figura 67).

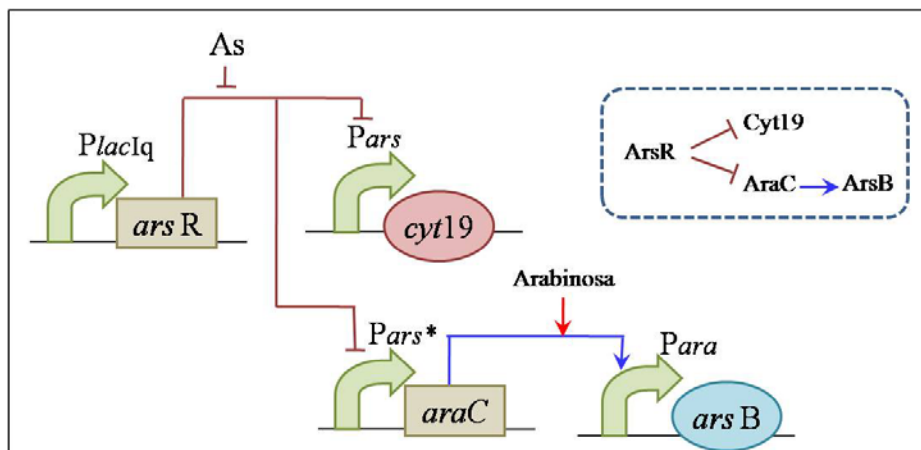


Figura 67. Configuración: mejora en el nivel de activación de cada mecanismo. En esta configuración se diferencia de la anterior en que tiene dos promotores diferentes, uno que expresa la metilación otro con el sistema de bombeo. El asterisco indica que las modificaciones para conseguir el nivel de activación necesario de la

bomba tendrían que dirigirse a la fuerza con la que se expresa el promotor en función de la represión que sufre.

Finalmente también sería deseable que en el momento en que es necesaria la activación de la bomba por ser la cantidad de arsénico muy elevada, ésta funcionara al máximo ya que el nivel de arsénico es peligroso para la supervivencia del organismo. Esto podría conseguirse con una nueva configuración donde tanto el sistema de metilación como el de bombeo están regulados por el nivel de arsénico (la metilasa se expresa bajo el promotor Pars y la bomba bajo un promotor modificado como en la configuración anterior), pero a su vez la bomba está activada por una proteína también controlada por el arsénico (está bajo el control de un promotor Pars modificado como el otro). De esta manera, en el momento que el nivel de arsénico sea suficiente como para eliminar la represión sobre la bomba y sobre la proteína activadora (en este caso AraC), ésta última produce una mayor activación en el sistema de bombeo. Esta configuración parece ser la más versátil de todas ya que consigue además de incrementar la pendiente del funcionamiento de la bomba, reducir el tiempo de respuesta de la bomba, y facilita la elección del nivel de activación para cada mecanismo (Figura 68).

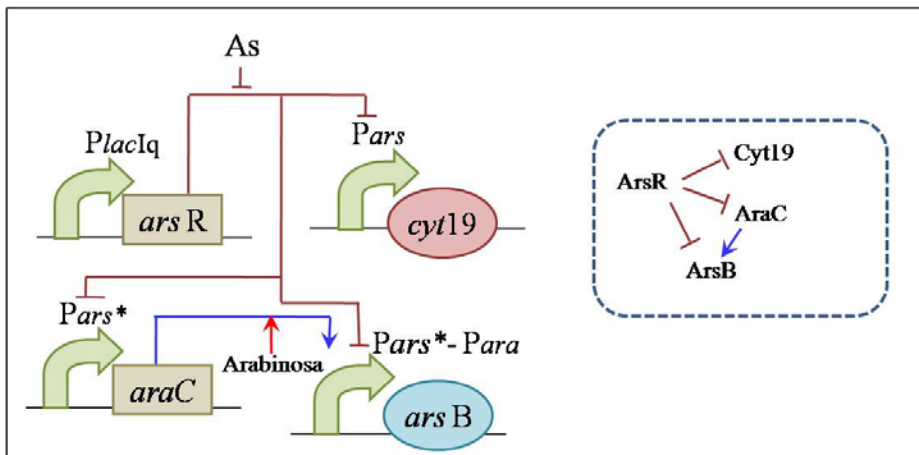


Figura 68. Configuración: mejora en el nivel de activación de cada mecanismo y en la respuesta del sistema de bombeo. En esta configuración cuando la cantidad de arsénico es suficiente para evitar la represión de la proteína ArsR sobre Pars* (promotor Pars modificado para que sea menos eficiente, es decir necesite más arsénico para que comience a expresarse) también se expresa el activador AraC incrementando la expresión de la bomba.

5 DISCUSIÓN

5.1 Aplicación de la Biología de sistemas en biodegradación: topología, organización y evolución.

El trabajo realizado en esta tesis se basa en el estudio de los procesos de biodegradación de compuestos tóxicos, realizados por microorganismos. Lo que diferencia a este trabajo de lo realizado anteriormente es la combinación de un análisis sistemático de la biodegradación, a diferencia de los estudios sobre rutas específicas (van der Meer *et al*, 1998; Arengi *et al*, 2001; Johnson & Spain, 2003; Shapir *et al*, 2007), la consideración de la biodegradación como un evento global independiente del organismo, a diferencia de los estudios de rutas en organismos específicos (Warhurst *et al*, 1994; Seeger *et al*, 1995; Casellas *et al*, 1997), la asignación manual de la metaenzima a las reacciones en lugar del uso del número enzimático EC y la utilización de una red de reacciones en lugar de una red de compuestos (Pazos *et al*, 2003).

Esta composición me ha permitido por una parte estructurar la información disponible en biodegradación de manera que sea posible su análisis computacional, y por otra parte obtener conocimiento global de su organización, formación y evolución, hasta el momento desconocido.

5.1.1 Base de datos BioNeMo.

El primer paso para cumplir con estos objetivos requirió la implementación de una base de datos relacional llamada BioNeMo junto con el desarrollo de un Interfaz de Programación de Aplicaciones (API) y el diseño y puesta en funcionamiento de un interfaz web. Los datos contenidos en BioNeMo suponen una mejora respecto a la información que se tenía de biodegradación debido a que he realizado la asignación manual de la entidad biológica a un gran número de reacciones bioquímicas y a la organización de estos datos en una estructura que permita el estudio de grandes cantidades de información y de sus relaciones. A su vez, el desarrollo de la API facilita enormemente el manejo de estos datos y permite su utilización para realizar estudios computacionales con esta información sin tener un

conocimiento exhaustivo de la estructura de la base de datos. Por último, el diseño y creación de la interfaz web amplía su utilización, ya que permite realizar consultas a la base de datos de forma muy sencilla, visual y sin necesidad de programar.

5.1.2 Análisis del nivel de descripción de la red de Biodegradación.

Para saber si los datos contenidos en BioNeMo eran adecuados para el estudio de la red de reacciones de biodegradación, cuantifiqué el nivel de descripción del sistema y su distribución dentro del mismo.

Ya que la red de biodegradación es una red metabólica al igual que la del Metabolismo Central, la cual ha sido ha sido extensivamente estudiada, tomé el nivel de caracterización sobre esta última como patrón para cuantificar el grado de descripción molecular de las reacciones en biodegradación.

La distribución de la información en el interior de la red la he obtenido clasificando las reacciones de biodegradación por una parte según su participación en procesos de degradación total o parcial, y por otra parte según su localización respecto al metabolismo central.

Los resultados obtenidos indican que la red de biodegradación ha sido mucho menos estudiada que la del Metabolismo Central y que las reacciones que participan en la biodegradación completa de los compuestos se han caracterizado mejor que las que no consiguen degradarlos totalmente hasta su entrada al CM. Sin embargo, la información que se tiene de la red de biodegradación está homogéneamente distribuida y por tanto aunque es necesario incrementar los esfuerzos para mejorar su caracterización, podemos utilizarla para identificar patrones o tendencias en su organización, formación, y evolución.

5.1.3 Topología de la red de Biodegradación.

El análisis de la topología ha demostrado que la red del Metabolismo Central y la red de compuestos de biodegradación comparten propiedades topológicas. La red de reacciones de biodegradación

presenta una estructura libre de escala al igual que las otras redes metabólicas. De este hecho se deduce que mientras que la mayoría de las reacciones tienen pocas conexiones, un número pequeño de ellas tiene una alta conectividad de entrada, salida o total.

De nuestro estudio podemos concluir que la red de reacciones de compuestos tóxicos (la red global de biodegradación) presenta una topología *ultra small-world* al igual que lo descrito para las otras redes metabólicas (Wagner & Fell, 2001; Pazos *et al*, 2003). Sin embargo a diferencia del comportamiento jerárquico observado en la red del Metabolismo Central, la red de reacciones de biodegradación no presenta jerarquía alguna. Es más, al igual que lo encontrado para la red de compuestos de biodegradación, la red de reacciones presenta una estructura muy lineal, introduciendo cierta interconexión a distancias próximas a la entrada al Metabolismo Central.

De los resultados anteriores (red *ultra small-world* pero no jerárquica) se deduce que, a pesar de que el diámetro de la red de reacciones de biodegradación es menor que el esperado para una red aleatoria análoga, comparado con la red del Metabolismo Central existe un incremento en el número de pasos necesarios para enlazar dos nodos de la red. Esto podría deberse a la carencia de nodos que conectan partes alejadas de la red teniendo que atravesar por lo tanto una cadena de reacciones poco conectadas.

Esta afirmación se ha visto reforzada al observar que la red de reacciones de biodegradación es muy lineal y solamente comienza a presentar más interconexión entre las reacciones en las cercanías del Metabolismo Central (al igual que lo observado en la red de compuestos de biodegradación). Como la mayor interconexión de la red se da cerca del Metabolismo Central y a esta distancia es donde se encuentran la mayoría de los nodos con alta conectividad, *hubs*, estos serían los responsables de producirla.

Relacionando el número de conexiones con el coeficiente de clustering, he encontrado que no existen nodos con una elevada conectividad y coeficiente de clustering al mismo tiempo. Esto indica que no hay candidatos a ser centro de clústeres de gran tamaño y que por lo tanto en la red de biodegradación, desde el punto de vista de reacciones, no existen grandes clústeres.

La ausencia de módulos y la abundancia de cadenas de reacciones lineales prácticamente carentes de interconexión excepto en la zona final de las rutas de biodegradación, confieren a la red una apariencia de embudo.

5.1.4 Organización de la red de Biodegradación.

Una vez conocida respecto a su topología, sus similitudes y diferencias con las otras redes metabólicas, he estudiado la organización de las reacciones dentro de la red, fijándome en las propiedades que tienen los complejos de proteínas que realizan cada reacción, y su variación dependiendo de su localización dentro de la red.

En la red de biodegradación la complejidad de las reacciones aumenta a medida que nos alejamos del Metabolismo Central. Esta complejidad es debida tanto al incremento del número de dominios por proteína como al aumento del número de proteínas que tienen los complejos. El concepto de complejidad indica la facilidad con la que los complejos podrían haber aparecido y/o evolucionado hacia una nueva función. De esta forma la interpretación de este resultado sería que los complejos que realizan las reacciones más alejadas (las de la periferia) les es más difícil evolucionar hacia nuevas funciones.

En general, en la red de biodegradación, las reacciones suelen ser específicas, (sus complejos tienen un alto grado de adaptación a la reacción que realizan). Sin embargo la distribución de ésta especificidad no es homogénea a lo largo de ella, disminuyendo a medida que nos alejamos del Metabolismo Central. Esto significa que los complejos más lejanos están menos adaptados a la función/actividad que realizan mientras que las próximas al metabolismo central están muy optimizadas para realizar una única reacción específica.

En general los complejos de las reacciones de biodegradación son homólogos entre sí. Solo he detectado dos casos donde las secuencias asociadas a una misma reacción no forman un clúster de complejos homólogos.

Respecto a la variabilidad de las proteínas de la red de biodegradación los resultados muestran que cuando se consideran solo los homólogos que conservan la especificidad de sustrato (la identidad de secuencia entre las proteínas es mayor del 95%), las proteínas alejadas del

Metabolismo Central tienen un mayor número de homólogos que las proteínas próximas al Metabolismo Central. Sin embargo cuando recupero proteínas homólogas simplemente fijándome que tengan la misma actividad catalítica (su identidad de secuencia sea superior al 50%), las proteínas a lo largo de toda la red presentan un número de proteínas homólogas parecido.

De los resultados se deduce que las proteínas alejadas del Metabolismo Central han divergido poco y mantienen la especificidad de sustrato, o mejor dicho la inespecificidad a sustrato ya que se ha visto que las reacciones en esta zona son poco específicas. Sin embargo la divergencia entre las proteínas cercanas al Metabolismo Central y sus homólogos es mucho mayor que para las proteínas alejadas del CM.

Una posible explicación a esta organización podría estar basada, por una parte, en que como se ha visto previamente, los complejos más simples pueden haber sido transferidos más fácilmente entre especies que los complejos multi-componente (Koonin, 2005), y por otra parte, en que los complejos cercanos al Metabolismo Central son los que aparecieron antes en la evolución (Pazos *et al*, 2003) y por lo tanto han tenido más tiempo para introducir mutaciones.

Los resultados obtenidos acerca de la organización de las reacciones dentro de la red de biodegradación y sus implicaciones en la adaptación, evolución y dispersión de las actividades, han ofrecido una primera aproximación hacia los mecanismos de evolución que subyacen en la red.

5.1.5 Evolución de la red de Biodegradación.

Del análisis funcional de la red de biodegradación se ha observado que el dominio funcional más “común” es el *2Fe-2S iron-sulfur cluster binding domain* (PF00111) y que la red de biodegradación se encuentra enriquecida en dominios relacionados con actividades de óxido-reducción. De la comparativa entre la frecuencia de los dominios propios de biodegradación en la red de biodegradación y en la red del Metabolismo Central, se ha obtenido que los dominios exclusivos de biodegradación se utilizan puntualmente y que los dominios más frecuentes en biodegradación también están presentes en el Metabolismo Central.

Terminada su comparación funcional, he analizado la similitud de secuencia entre las proteínas de la red de biodegradación y la red del Metabolismo Central. Los resultados obtenidos indican que solo un porcentaje muy pequeño de las proteínas en biodegradación (el 25%) tienen una relación evolutiva identificable por homología de secuencia con alguna proteína del Metabolismo Central. El 75% de las proteínas restantes tampoco tenían una relación evolutiva detectable en otros sistemas biológicos a excepción de la propia red de biodegradación, donde el 55% de ellos tienen un homólogo cercano. Aunque aparentemente contradictorio, este resultado se puede conciliar con el obtenido en la comparación funcional de ambas redes mediante dos posibles explicaciones. La primera es que aunque los dominios de las proteínas conocidas en biodegradación ya existían en el Metabolismo Central, su ensamblaje en proteínas es diferente. La segunda posibilidad es que las proteínas de biodegradación, además de tener dichos dominios tengan otros no conocidos y que no se encuentran en las proteínas del Metabolismo Central.

Existen dos fenómenos (no excluyente entre sí) que podrían explicar porque no he detectado relación evolutiva entre las proteínas de biodegradación y del Metabolismo Central. Una es que las proteínas de la red de biodegradación vengan del Metabolismo Central pero que hayan evolucionado rápidamente. La otra es que las proteínas de biodegradación no vengan del Metabolismo Central y procedan de un pequeño subconjunto de complejo-enzimas encargados del metabolismo primigenio de compuestos naturales similares a los recalcitrantes, como son el fenol y el tolueno (Janssen *et al*, 2005).

A nivel de reacciones, los resultados sugieren que la mayoría de las reacciones cuentan con un complejo homogéneo “CM” o “no CM”. Los casos de reacciones mixtas son muy escasos y se encuentran alejados de la entrada al Metabolismo Central. Aunque no es posible afirmarlo debido a la pobre caracterización de estos complejos mixtos, parece que de su relación evolutiva con el Metabolismo Central se podría haber obtenido la actividad catalítica. Además al ser en la zona en la que aparecen las reacciones mixtas los compuestos más complejos y diferentes de los utilizados en el Metabolismo Central (Pazos *et al*, 2003), se puede pensar que las proteínas que no tienen una relación evolutiva detectable con el Metabolismo Central podrían estar implicadas en la unión a substrato y a cofactor.

También he visto que casi la totalidad de los pares de reacciones en biodegradación que siendo consecutivos mantienen una relación evolutiva detectable con reacciones del Metabolismo Central, han sido reclutadas en bloque (las reacciones del Metabolismo Central también son consecutivas) apareciendo en las proximidades del Metabolismo Central. Esto podría deberse a que en esta zona se realizan actividades parecidas y se degradan compuestos similares (Pazos *et al*, 2003) a los que se dan en la red del Metabolismo Central.

Teniendo en cuenta los mecanismos evolutivos que se ha descrito que participan en la evolución de la red del Metabolismo Central (Horowitz, 1945; Horowitz, 1965; Ycas, 1974; Jensen, 1976; Huynen & Snel, 2000; Rison & Thornton, 2002): modelo de retro-evolución, de duplicación de rutas y de mosaico, se identificó y cuantificó su aportación en el proceso de formación y evolución de la red de biodegradación. Los resultados obtenidos muestran que el mecanismo más utilizado es el de mosaico, seguido del modelo de duplicación de rutas y en raras ocasiones el modelo de retro-evolución. Este orden de frecuencia de aparición coincide con el observado anteriormente en el Metabolismo Central. Por otra parte, estos mecanismos de evolución no se distribuyen homogéneamente a lo largo de la red de biodegradación estando más representado el modelo de retro-evolución en las proximidades del Metabolismo Central y el modelo de mosaico y de duplicación de rutas en zonas más alejadas.

La fuerza que guía la formación de una ruta metabólica que sigue el modelo de retro-evolución es el intento de conseguir mediante la transformación de un precursor “A” un metabolito “X” que el organismo ha agotado y necesita. Una vez que la ruta ha incorporado este precursor “A”, al utilizarlo llegará un momento que lo agote y por lo tanto necesite buscar un nuevo precursor “B” a partir del cual produzca el precursor “A”, que producirá a su vez el metabolito “X” que le era necesario. Por este motivo se dice que el modelo de retro-evolución está principalmente dirigido a la obtención de un producto final. Por otra parte en las rutas formadas siguiendo un modelo de mosaico o de duplicación de rutas el objetivo es utilizar mecanismos catalíticos existentes (y que realicen transformaciones similares) para ir cubriendo las reacciones necesarias en la sucesión de los pasos metabólicos de la nueva ruta. Por ello se dice que estos eventos están dirigidos a la innovación y diversificación de las actividades enzimáticas (Schmidt *et al*, 2003). Por lo tanto, la distribución de los mecanismos en la red de biodegradación podría estar relacionada con la

presión selectiva diferente que existe a lo largo del proceso de biodegradación. En los primeros pasos de la red, al ser los compuestos muy tóxicos, el objetivo es degradar los compuestos, transformarlos para poder sobrevivir. Sin embargo en los últimos pasos de la red de biodegradación, los compuestos han perdido toxicidad y por lo tanto ahora el objetivo es la obtención de un producto final concreto que proporcione energía o nutrientes, es decir, que se enlace con la Red del Metabolismo Central.

Por último, he mostrado que en eventos de retro-evolución y de mosaico la duplicación genética es más abundante, el par de RRSS son realizadas por complejos de proteínas que se habrían duplicado y posteriormente divergido. Sin embargo en eventos de duplicación de rutas la dualidad funcional es más frecuente, el par de RRSS son realizadas por el mismo complejo de proteínas.

La dualidad funcional implica que las proteínas son capaces de realizar reacciones secundarias, que aun no siendo su “especialidad” y por lo tanto no realizándolas de manera optimizada, lo hacen a un nivel de eficiencia que puede ser potencialmente útil (Copley, 2000). Sin embargo la duplicación genética conlleva que cuando un gen se duplica, la copia del gen es libre de mutar. Gracias a estas mutaciones, la proteína codificada por este gen realizará la nueva función de manera más eficiente de lo que lo puede realizarla la proteína del gen inicial. Sin embargo para esto se necesita tiempo, el cuál variará dependiendo de la tasa de mutación del gen. Por lo tanto, lo habitual en los eventos de retro-evolución y de mosaico es que las secuencias de los complejos de proteínas de las reacciones que se comparan diverjan en un cierto número de mutaciones que les permite adecuarse mejor a la nueva reacción. Sin embargo en la mayoría de los eventos de duplicación de rutas, un mismo complejo de proteínas realiza transformaciones diferentes con mayor o menor eficiencia.

5.2 Teoría sobre la formación y evolución de la red de biodegradación.

Teniendo en cuenta toda esta información podemos concluir que la red de biodegradación es muy lineal y poco interconectada lo que indicaría que la manera en la que la red degrada un nuevo compuesto es o añadiendo pasos en los extremos de la red (creciendo verticalmente) cuando se trata de compuestos más complejos de los existentes, o añadiendo nuevas rutas completas y lineales. Este fenómeno explicaría porque la red tiene forma de embudo y porque a pesar de ser *ultra-small world* presenta una longitud de caminos superior al de la red del Metabolismo Central.

Basándome en el trabajo realizado por Pazos donde se vio que las enzimas más antiguas eran las cercanas al metabolismo central ((Pazos *et al*, 2003)) y tras analizar mis resultados acerca de la organización de las reacciones en la red podemos enunciar que los complejos de proteínas que se encuentran cerca del punto de entrada al Metabolismo Central son los más antiguos de la red. Al haber tenido tiempo y al ser más sencillos, son capaces de evolucionar más fácilmente y así adaptarse mejor a la actividad que realizan y al sustrato que degradan. Sin embargo los complejos de proteínas que se encuentran en la periferia de la red, independientemente de que sean más o menos antiguos, al ser más complejos (tener más dominios), su evolución hacia una nueva función es menos probable. La manera de solventar este problema radica en la poca adaptación de estos complejos de proteínas a una actividad específica y su poca especificidad de sustrato, lo cual permite degradar nuevos compuestos. Por tanto, la inespecificidad de los complejos es una ventaja a la hora de detoxificar. Esta teoría corroboraría lo anteriormente dicho sobre que la red de biodegradación crece verticalmente.

Las funciones de la red de biodegradación no son muy diferentes de las que existen en el Metabolismo Central, sin embargo la mayoría de sus proteínas y reacciones (la maquinaria enzimática) no tiene relación evolutiva identificable por comparación de secuencias con la red del Metabolismo Central. Este hecho conduce a que la formación de la red de biodegradación se dé a partir de las rutas de degradación de compuestos naturales estructuralmente análogos donde se realizan actividades similares y que su evolución sea el producto de la acción de

mecanismos internos, siendo él más representativo él de mosaico seguido por él de duplicación de rutas.

Hemos mencionado que la presión selectiva en diferentes zonas de la red es distinta, y que esto produce que los mecanismos de evolución se distribuyan de una cierta manera. Sin embargo, esta presión selectiva puede que también explique la abundancia de la duplicación genética o la dualidad funcional en cada mecanismo evolutivo.

Mi interpretación es la siguiente. En la zona externa de la red es necesario dar una respuesta rápida para comenzar la biodegradación debido a la toxicidad de los compuestos. Esto se traduce en la falta de tiempo para que se produzcan mutaciones y por ello en esa zona de la red abundan los eventos de duplicación de rutas con dualidad funcional (los cuales producen inespecificidad). Sin embargo cuando nos encontramos cerca del metabolismo central, la toxicidad de los compuestos disminuye, y por tanto no hay una presión a dar una respuesta rápida, sino a optimizar las transformaciones para conseguir energía o nutrientes. Por ello, los casos de retro-evolución por duplicación genética son más habituales en esta región.

6 CONCLUSIONES

1. La red de reacciones de biodegradación presenta una estructura libre de escala y *ultra-small world* al igual que lo descrito para la red del Metabolismo Central. Sin embargo a diferencia de esta última red, no tiene una estructura jerárquica siendo sus caminos muy lineales e interconectados en las proximidades a la entrada al Metabolismo Central, dotando a la red de biodegradación con una estructura de embudo.
2. Conforme nos alejamos de la entrada al Metabolismo Central la especificidad de los complejos de proteínas disminuye y su complejidad aumenta. Además los grupos de proteínas homólogas en las zonas alejadas del Metabolismo Central tienen menos diversidad de secuencia (las proteínas se parecen mucho entre sí).
3. Salvo unos pocos dominios funcionales propios de biodegradación, los dominios de las proteínas de la red de biodegradación, también se encuentran en proteínas del Metabolismo Central. Sin embargo no he detectado relación evolutiva por similitud de secuencia entre la mayoría de las proteínas de biodegradación y las del Metabolismo Central. Por tanto, los dominios se ensamblan con dominios no conocidos o de manera diferente en las proteínas de la red de biodegradación
4. La mayoría de las reacciones tienen complejos de proteínas con origen homogéneo (o todas sus proteínas o ninguna tiene relación evolutiva detectable por similitud de secuencia con el Metabolismo Central). Los complejos con origen mixto parecen haber tomado al menos la actividad catalítica del Metabolismo Central. Además si dos reacciones importadas del Metabolismo Central eran consecutivas en éste, en la red de biodegradación lo siguen siendo (se han importado juntas).
5. Al igual que lo descrito para la red del Metabolismo Central el mecanismo evolutivo más frecuente es el llamado mosaico, seguido de la duplicación de rutas y muy por debajo la retro-evolución. Los eventos de retro-evolución, aunque minoritarios a cualquier distancia, son más habituales en las proximidades del Metabolismo Central. Sin embargo los eventos de mosaico y duplicación de rutas son más frecuentes en la zona externa de la red de biodegradación. Existe una tendencia a encontrar duplicación genéticas en los mecanismos de

mosaico y de retro-evolución y dualidad funcional en los eventos de duplicación de rutas.

6. Es posible, haciendo uso de la biología sintética, diseñar circuitos capaces de contribuir a la eliminación de contaminantes ambientales. Mediante el modelado y simulación de circuitos biológicos se puede verificar el correcto funcionamiento del sistema, comprobar tiempos de respuesta e indicar a que parámetros hay que prestar un especial interés en la implementación experimental.

7 BIBLIOGRAFÍA

(2007) The Universal Protein Resource (UniProt). *Nucleic Acids Res* **35**(Database issue): D193-197

Abraham WR, Nogales B, Golyshin PN, Pieper DH, Timmis KN (2002) Polychlorinated biphenyl-degrading microbial communities in soils and sediments. *Curr Opin Microbiol* **5**(3): 246-253

Albert R, Jeong H, Barabasi AL (2000) Error and attack tolerance of complex networks. *Nature* **406**(6794): 378-382

Alexander M (2001) *Biodegradation and Bioremediation*: ACADEMIC PRESS.

Alon U (2003) Biological networks: the tinkerer as an engineer. *Science* **301**(5641): 1866-1867

Alves R, Chaleil RA, Sternberg MJ (2002) Evolution of enzymes in metabolism: a network perspective. *J Mol Biol* **320**(4): 751-770

Arengi FL, Berlanda D, Galli E, Sello G, Barbieri P (2001) Organization and regulation of meta cleavage pathway genes for toluene and o-xylene derivative degradation in *Pseudomonas stutzeri* OX1. *Appl Environ Microbiol* **67**(7): 3304-3308

Atlas RMU, R. (1999) Bioremediation. In *Manual of Industrial Microbiology and Biotechnology 2nd edition*, J.E. DALD (ed), pp 666 - 681. Washington D.C.: ASM Press

Barabasi AL, Albert R (1999) Emergence of scaling in random networks. *Science* **286**(5439): 509-512

Barabasi AL, Bonabeau E (2003) Scale-free networks. *Sci Am* **288**(5): 60-69

Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5**(2): 101-113

Barkay T, Pritchard H (1988) Adaptation of aquatic microbial communities to pollutant stress. *Microbiol Sci* **5**(6): 165-169

Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R (2005) A synthetic multicellular system for programmed pattern formation. *Nature* **434**(7037): 1130-1134

Basu S, Mehreja R, Thiberge S, Chen MT, Weiss R (2004) Spatiotemporal control of gene expression with pulse-generating networks. *Proc Natl Acad Sci U S A* **101**(17): 6355-6360

Beard J (2006) DDT and human health. *Sci Total Environ* **355**(1-3): 78-89

Becskei A, Seraphin B, Serrano L (2001) Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J* **20**(10): 2528-2535

Becskei A, Serrano L (2000) Engineering stability in gene networks by autoregulation. *Nature* **405**(6786): 590-593

Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C (2002) The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* **58**(Pt 6 No 1): 899-907

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* **28**(1): 235-242

Bio FABG, Baker D, Church G, Collins J, Endy D, Jacobson J, Keasling J, Modrich P, Smolke C, Weiss R (2006) Engineering life: building a fab for biology. *Sci Am* **294**(6): 44-51

Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* **31**(1): 365-370

Bork P, Sander C, Valencia A (1993) Convergent evolution of similar enzymatic function on different protein folds: the hexokinase, ribokinase, and galactokinase families of sugar kinases. *Protein Sci* **2**(1): 31-40

Bouwer EJ, Zehnder AJ (1993) Bioremediation of organic compounds-putting microbial metabolism to work. *Trends Biotechnol* **11**(8): 360-367

Brokamp A, Happe B, Schmidt FR (1996) Cloning and nucleotide sequence of a D,L-haloalkanoic acid dehalogenase encoding gene from *Alcaligenes xylosoxidans* ssp. *denitrificans* ABIV. *Biodegradation* **7**(5): 383-396

Canovas D, Cases I, de Lorenzo V (2003) Heavy metal tolerance and metal homeostasis in *Pseudomonas putida* as revealed by complete genome analysis. *Environ Microbiol* **5**(12): 1242-1256

Carrera G, Fernandez P, Grimalt JO, Ventura M, Camarero L, Catalan J, Nickus U, Thies H, Psenner R (2002) Atmospheric deposition of organochlorine compounds to remote high mountain lakes of Europe. *Environ Sci Technol* **36**(12): 2581-2588

Casellas M, Grifoll M, Bayona JM, Solanas AM (1997) New metabolites in the degradation of fluorene by *Arthrobacter* sp. strain F101. *Appl Environ Microbiol* **63**(3): 819-826

Cases I, de Lorenzo V (2005) Genetically modified organisms for the environment: stories of success and failure and what we have learned from them. *Int Microbiol* **8**(3): 213-222

Caspi R, Foerster H, Fulcher CA, Kaipa P, Krummenacker M, Latendresse M, Paley S, Rhee SY, Shearer AG, Tissier C, Walk TC, Zhang P, Karp PD (2007) The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res*

Copley SD (2000) Evolution of a metabolic pathway for degradation of a toxic xenobiotic: the patchwork approach. *Trends Biochem Sci* **25**(6): 261-265

Chakrabarty AM, Chou G, Gunsalus IC (1973) Genetic regulation of octane dissimilation plasmid in *Pseudomonas*. *Proc Natl Acad Sci U S A* **70**(4): 1137-1140

Chen CM, Misra TK, Silver S, Rosen BP (1986) Nucleotide sequence of the structural genes for an anion pump. The plasmid-encoded arsenical resistance operon. *J Biol Chem* **261**(32): 15030-15038

Chen MT, Weiss R (2005) Artificial cell-cell communication in yeast *Saccharomyces cerevisiae* using signaling elements from *Arabidopsis thaliana*. *Nat Biotechnol* **23**(12): 1551-1555

de Lorenzo V (2001) Cleaning up behind us. The potential of genetically modified bacteria to break down toxic pollutants in the environment. *EMBO Rep* **2**(5): 357-359

Dejonghe W, Goris J, El Fantroussi S, Hofte M, De Vos P, Verstraete W, Top EM (2000) Effect of dissemination of 2,4-dichlorophenoxyacetic acid (2,4-D) degradation plasmids on 2,4-D degradation and on bacterial community structure in two different soil horizons. *Appl Environ Microbiol* **66**(8): 3297-3304

Devos D, Valencia A (2000) Practical limits of function prediction. *Proteins* **41**(1): 98-107

Devos D, Valencia A (2001) Intrinsic errors in genome annotation. *Trends Genet* **17**(8): 429-431

Diaz E (2004) Bacterial degradation of aromatic pollutants: a paradigm of metabolic versatility. *Int Microbiol* **7**(3): 173-180

Diaz E, Ferrandez A, Garcia JL (1998) Characterization of the hca cluster encoding the dioxygenolytic pathway for initial catabolism of 3-phenylpropionic acid in *Escherichia coli* K-12. *J Bacteriol* **180**(11): 2915-2923

Diaz E, Munthali M, de Lorenzo V, Timmis KN (1994) Universal barrier to lateral spread of specific genes among microorganisms. *Mol Microbiol* **13**(5): 855-861

Doolittle RF (1994) Convergent evolution: the need to be explicit. *Trends Biochem Sci* **19**(1): 15-18

Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* **403**(6767): 335-338

Ellis LB, Roe D, Wackett LP (2006) The University of Minnesota Biocatalysis/Biodegradation Database: the first decade. *Nucleic Acids Res* **34**(Database issue): D517-521

Endy D (2005) Foundations for engineering biology. *Nature* **438**(7067): 449-453

Englesberg E, Irr J, Power J, Lee N (1965) Positive control of enzyme synthesis by gene C in the L-arabinose system. *J Bacteriol* **90**(4): 946-957

Feng XJ, Hooshangi S, Chen D, Li G, Weiss R, Rabitz H (2004) Optimizing genetic circuits by global sensitivity analysis. *Biophys J* **87**(4): 2195-2202

Finn RD, Tate J, Mistry J, Coghill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A (2007) The Pfam protein families database. *Nucleic Acids Res*

Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**(6767): 339-342

Gerlt JA, Babbitt PC (2001) Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu Rev Biochem* **70**: 209-246

Gill SR, Fouts DE, Archer GL, Mongodin EF, Deboy RT, Ravel J, Paulsen IT, Kolonay JF, Brinkac L, Beanan M, Dodson RJ, Daugherty SC, Madupu R, Angiuoli SV, Durkin AS, Haft DH, Vamathevan J, Khouri H, Utterback T, Lee C, Dimitrov G, Jiang L, Qin H, Weidman J, Tran K, Kang K, Hance IR, Nelson KE, Fraser CM (2005) Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus epidermidis* strain. *J Bacteriol* **187**(7): 2426-2438

Gomez MJ, Pazos F, Guijarro FJ, de Lorenzo V, Valencia A (2007) The environmental fate of organic pollutants through the global microbial metabolism. *Mol Syst Biol* **3**: 114

Greenblatt J, Schleif R (1971) Arabinose C protein: regulation of the arabinose operon in vitro. *Nat New Biol* **233**(40): 166-170

- Hartwell LH, Hopfield JJ, Leibler S, Murray AW (1999) From molecular to modular cell biology. *Nature* **402**(6761 Suppl): C47-52
- Hatzimanikatis V, Li C, Ionita JA, Broadbelt LJ (2004) Metabolic networks: enzyme function and metabolite structure. *Curr Opin Struct Biol* **14**(3): 300-306
- Hooshangi S, Thiberge S, Weiss R (2005) Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc Natl Acad Sci U S A* **102**(10): 3581-3586
- Horowitz NH (1945) On the Evolution of Biochemical Syntheses. *Proc Natl Acad Sci U S A* **31**(6): 153-157
- Horowitz NH (1965) The evolution of biochemical synthesis: retrospect and prospect. In *Evolving genes and proteins*, Brison VV, H. (ed), pp 15 - 23. New York: Academic Press
- Huynen MA, Snel B (2000) Gene and context: integrative approaches to genome analysis. *Adv Protein Chem* **54**: 345-379
- Jacob F, Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* **3**: 318-356
- Janssen DB, Dinkla IJ, Poelarends GJ, Terpstra P (2005) Bacterial degradation of xenobiotic compounds: evolution and distribution of novel enzyme activities. *Environ Microbiol* **7**(12): 1868-1882
- Jensen RA (1976) Enzyme recruitment in evolution of new function. *Annu Rev Microbiol* **30**: 409-425
- Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL (2000) The large-scale organization of metabolic networks. *Nature* **407**(6804): 651-654
- Johnson GR, Spain JC (2003) Evolution of catabolic pathways for synthetic compounds: bacterial pathways for degradation of 2,4-dinitrotoluene and nitrobenzene. *Appl Microbiol Biotechnol* **62**(2-3): 110-123
- Jones FT (2007) A broad view of arsenic. *Poult Sci* **86**(1): 2-14

Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**(1): 27-30

Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* **34**(Database issue): D354-357

Karp PD, Riley M, Paley SM, Pellegrini-Toole A (2002) The MetaCyc Database. *Nucleic Acids Res* **30**(1): 59-61

King RB (1997) *Practical environmental bioremediation, the field guide*: Lewis Publishers (CRC Press).

Kitano H (2001) *Foundations of Systems Biology*: MIT Press.

Koonin EV (2005) Systemic determinants of gene evolution and function. *Mol Syst Biol* **1**: 2005 0021

Lazcano A, Miller SL (1999) On the origin of metabolic pathways. *J Mol Evol* **49**(4): 424-431

Lin S, Shi Q, Nix FB, Styblo M, Beck MA, Herbin-Davis KM, Hall LL, Simeonsson JB, Thomas DJ (2002) A novel S-adenosyl-L-methionine:arsenic(III) methyltransferase from rat liver cytosol. *J Biol Chem* **277**(13): 10795-10803

Lindow SE, Panopoulos NJ, McFarland BL (1989) Genetic engineering of bacteria from managed and natural habitats. *Science* **244**(4910): 1300-1307

Liu S, Suflita JM (1993) Ecology and evolution of microbial populations for bioremediation. *Trends Biotechnol* **11**(8): 344-352

Looger LL, Dwyer MA, Smith JJ, Hellinga HW (2003) Computational design of receptor and sensor proteins with novel functions. *Nature* **423**(6936): 185-190

Maglott D, Ostell J, Pruitt KD, Tatusova T (2007) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res* **35**(Database issue): D26-31

Minshull J (1995) Cleaning up our own backyard: developing new catabolic pathways to degrade pollutants. *Chem Biol* **2**(12): 775-780

Mishra V, Lal R, Srinivasan (2001) Enzymes and operons mediating xenobiotic degradation in bacteria. *Crit Rev Microbiol* **27**(2): 133-166

Neidle EL, Hartnett C, Ornston LN, Bairoch A, Rekik M, Harayama S (1991) Nucleotide sequences of the *Acinetobacter calcoaceticus* benABC genes for benzoate 1,2-dioxygenase reveal evolutionary relationships among multicomponent oxygenases. *J Bacteriol* **173**(17): 5385-5395

Neyt C, Iriarte M, Thi VH, Cornelis GR (1997) Virulence and arsenic resistance in *Yersinia*. *J Bacteriol* **179**(3): 612-619

Noble D (2006) *The Music of Life*: Oxford University Press.

Parter M, Kashtan N, Alon U (2007) Environmental variability and modularity of bacterial metabolic networks. *BMC Evol Biol* **7**: 169

Pazos F, Guijas D, Valencia A, De Lorenzo V (2005) MetaRouter: bioinformatics for bioremediation. *Nucleic Acids Res* **33**(Database issue): D588-592

Pazos F, Valencia A, De Lorenzo V (2003) The organization of the microbial biodegradation network from a systems-biology perspective. *EMBO Rep* **4**(10): 994-999

Pelz O, Tesar M, Wittich RM, Moore ER, Timmis KN, Abraham WR (1999) Towards elucidation of microbial community metabolic pathways: unravelling the network of carbon sharing in a pollutant-degrading bacterial consortium by immunocapture and isotopic ratio mass spectrometry. *Environ Microbiol* **1**(2): 167-174

Petsko GA, Kenyon GL, Gerlt JA, Ringe D, Kozarich JW (1993) On the origin of enzymatic species. *Trends Biochem Sci* **18**(10): 372-376

Pruitt KD, Tatusova T, Maglott DR (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **35**(Database issue): D61-65

Ptashne M (2004) *A Genetic Switch*, New York: Cold Spring Harbor Laboratory Press.

Radhika V, Proikas-Cezanne T, Jayaraman M, Onesime D, Ha JH, Dhanasekaran DN (2007) Chemical sensing of DNT by engineered olfactory yeast strain. *Nat Chem Biol* **3**(6): 325-330

Ramos JL, Diaz E, Dowling D, de Lorenzo V, Molin S, O'Gara F, Ramos C, Timmis KN (1994) The behavior of bacteria designed for biodegradation. *Biotechnology (N Y)* **12**(13): 1349-1356

Ramos JL, Mermod N, Timmis KN (1987) Regulatory circuits controlling transcription of TOL plasmid operon encoding meta-cleavage pathway for degradation of alkylbenzoates by *Pseudomonas*. *Mol Microbiol* **1**(3): 293-300

Ravasz E, Barabasi AL (2003) Hierarchical organization in complex networks. *Phys Rev E Stat Nonlin Soft Matter Phys* **67**(2 Pt 2): 026112

Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL (2002) Hierarchical organization of modularity in metabolic networks. *Science* **297**(5586): 1551-1555

Reineke W (1998) Development of hybrid strains for the mineralization of chloroaromatics by patchwork assembly. *Annu Rev Microbiol* **52**: 287-331

Rison SC, Thornton JM (2002) Pathway evolution, structurally speaking. *Curr Opin Struct Biol* **12**(3): 374-382

Rojo F, Pieper DH, Engesser KH, Knackmuss HJ, Timmis KN (1987) Assemblage of ortho cleavage route for simultaneous degradation of chloro- and methylaromatics. *Science* **238**(4832): 1395-1398

Rosenberg ER, E.Z. (1996) Bioremediation of petroleum contamination. In *Bioremediation. Principles and Applications*, DL CRC (ed), 4, pp 100 - 124. Cambridge University Press

Sauer U, Heinemann M, Zamboni N (2007) Genetics. Getting closer to the whole picture. *Science* **316**(5824): 550-551

Sayler GS, Ripp S (2000) Field applications of genetically engineered microorganisms for bioremediation processes. *Curr Opin Biotechnol* **11**(3): 286-289

Schmidt S, Sunyaev S, Bork P, Dandekar T (2003) Metabolites: a helping hand for pathway evolution? *Trends Biochem Sci* **28**(6): 336-341

Seeger M, Timmis KN, Hofer B (1995) Degradation of chlorobiphenyls catalyzed by the bph-encoded biphenyl-2,3-dioxygenase and biphenyl-2,3-dihydrodiol-2,3-dehydrogenase of *Pseudomonas* sp. LB400. *FEMS Microbiol Lett* **133**(3): 259-264

Shaham M, Chakrabarty AM, Gunsalus IC (1973) Camphor plasmid-mediated chromosomal transfer in *Pseudomonas putida*. *J Bacteriol* **116**(2): 944-949

Shapir N, Mongodin EF, Sadowsky MJ, Daugherty SC, Nelson KE, Wackett LP (2007) Evolution of catabolic pathways: Genomic insights into microbial s-triazine metabolism. *J Bacteriol* **189**(3): 674-682

Springael D, Top EM (2004) Horizontal gene transfer and microbial adaptation to xenobiotics: new types of mobile genetic elements and lessons from ecological studies. *Trends Microbiol* **12**(2): 53-58

Strong LC, McTavish H, Sadowsky MJ, Wackett LP (2000) Field-scale remediation of atrazine-contaminated soil using recombinant *Escherichia coli* expressing atrazine chlorohydrolase. *Environ Microbiol* **2**(1): 91-98

Suris-Regueiro JC, Garza-Gil MD, Varela-Lafuente MM (2007) The Prestige oil spill and its economic impact on the Galician fishing sector. *Disasters* **31**(2): 201-215

Tamames J, Moya A, Valencia A (2007) Modular organization in the reductive evolution of protein-protein interaction networks. *Genome Biol* **8**(5): R94

Teichmann SA, Rison SC, Thornton JM, Riley M, Gough J, Chothia C (2001a) The evolution and structural anatomy of the small molecule metabolic pathways in *Escherichia coli*. *J Mol Biol* **311**(4): 693-708

Teichmann SA, Rison SC, Thornton JM, Riley M, Gough J, Chothia C (2001b) Small-molecule metabolism: an enzyme mosaic. *Trends Biotechnol* **19**(12): 482-486

Tian W, Skolnick J (2003) How well is enzyme function conserved as a function of pairwise sequence identity? *J Mol Biol* **333**(4): 863-882

Timmis KN, Pieper DH (1999) Bacteria designed for bioremediation. *Trends Biotechnol* **17**(5): 200-204

Top EM, Springael D (2003) The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. *Curr Opin Biotechnol* **14**(3): 262-269

Travers JMS (1969) An Experimental Study of the Small World Problem. *Sociometry* **32**(4): 425-443

Tucker JB, Zilinskas RA (2006) The promise and perils of synthetic biology. *New Atlantis* **12**: 25-45

van der Meer JR, de Vos WM, Harayama S, Zehnder AJ (1992) Molecular mechanisms of genetic adaptation to xenobiotic compounds. *Microbiol Rev* **56**(4): 677-694

van der Meer JR, Werlen C, Nishino SF, Spain JC (1998) Evolution of a pathway for chlorobenzene metabolism leads to natural attenuation in contaminated groundwater. *Appl Environ Microbiol* **64**(11): 4185-4193

van der Ploeg J, van Hall G, Janssen DB (1991) Characterization of the haloacid dehalogenase from *Xanthobacter autotrophicus* GJ10 and sequencing of the *dhlB* gene. *J Bacteriol* **173**(24): 7925-7933

von Canstein H, Li Y, Timmis KN, Deckwer WD, Wagner-Dobler I (1999) Removal of mercury from chloralkali electrolysis wastewater by a mercury-resistant *Pseudomonas putida* strain. *Appl Environ Microbiol* **65**(12): 5279-5284

Wackett LP (2004) Evolution of enzymes for the metabolism of new chemical inputs into the environment. *J Biol Chem* **279**(40): 41259-41262

Wagner A, Fell DA (2001) The small world inside large metabolic networks. *Proc Biol Sci* **268**(1478): 1803-1810

Walter MV (1997) Bioaugmentation. In *Manual of environmental microbiology*, C.J. H (ed), pp 753–765. Washington, DC: ASM Press

Walton FS, Waters SB, Jolley SL, LeCluyse EL, Thomas DJ, Styblo M (2003) Selenium compounds modulate the activity of recombinant rat AsIII-methyltransferase and the methylation of arsenite by rat and human hepatocytes. *Chem Res Toxicol* **16**(3): 261-265

Warhurst AM, Clarke KF, Hill RA, Holt RA, Fewson CA (1994) Metabolism of styrene by *Rhodococcus rhodochrous* NCIMB 13259. *Appl Environ Microbiol* **60**(4): 1137-1145

Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. *Nature* **393**(6684): 440-442

Werlen C, Kohler HP, van der Meer JR (1996) The broad substrate chlorobenzene dioxygenase and cis-chlorobenzene dihydrodiol dehydrogenase of *Pseudomonas* sp. strain P51 are linked evolutionarily to the enzymes for benzene and toluene degradation. *J Biol Chem* **271**(8): 4009-4016

Wilkins BM (2002) Plasmid promiscuity: meeting the challenge of DNA immigration control. *Environ Microbiol* **4**(9): 495-500

Wüthrich K (1986) *NMR of Proteins and Nucleic Acids*, New Jersey: Wiley-Interscience.

Ycas M (1974) On earlier states of the biochemical system. *J Theor Biol* **44**(1): 145-160

8 Apéndice 1

Carbohydrate Metabolism

- Glycolysis / Gluconeogenesis (map00010)
- Citrate cycle (TCA cycle) (map00020)
- Pentose phosphate pathway (map00030)
- Pentose and glucuronate interconversions (map00040)
- Fructose and mannose metabolism (map00051)
- Galactose metabolism (map00052)
- Ascorbate and aldarate metabolism (map00053)
- Starch and sucrose metabolism (map00500)
- Aminosugars metabolism (map00530)
- Nucleotide sugars metabolism (map00520)
- Pyruvate metabolism (map00620)
- Glyoxylate and dicarboxylate metabolism (map00630)
- Propanoate metabolism (map00640)
- Butanoate metabolism (map00650)
- Branched dibasic acid metabolism (map00660)
- Inositol metabolism (map00031)
- Inositol phosphate metabolism (map00562)

Energy Metabolism

- Oxidative phosphorylation (map00190)
- ATP synthesis (map00193)
- Photosynthesis (map00195)
- Carbon fixation (map00710)
- Reductive carboxylate cycle (CO₂ fixation) (map00720)
- Methane metabolism (map00680)
- Nitrogen metabolism (map00910)
- Sulfur metabolism (map00920)

Lipid Metabolism

- Fatty acid biosynthesis (map00061)
- Fatty acid elongation in mitochondria (map00062)
- Fatty acid metabolism (map00071)
- Synthesis and degradation of ketone bodies (map00072)
- Biosynthesis of steroids (map00100)
- Bile acid biosynthesis (map00120)
- C21-Steroid hormone metabolism (map00140)
- Androgen and estrogen metabolism (map00150)

Glycerolipid metabolism (map00561)
Glycerophospholipid metabolism (map00564)
Prostaglandin and leukotriene metabolism (map00590)

Nucleotide Metabolism

Purine metabolism (map00230)
Pyrimidine metabolism (map00240)

Amino Acid Metabolism

Glutamate metabolism (map00251)
Alanine and aspartate metabolism (map00252)
Glycine, serine and threonine metabolism (map00260)
Methionine metabolism (map00271)
Cysteine metabolism (map00272)
Valine, leucine and isoleucine degradation (map00280)
Valine, leucine and isoleucine biosynthesis (map00290)
Lysine biosynthesis (map00300)
Lysine degradation (map00310)
Arginine and proline metabolism (map00330)
Histidine metabolism (map00340)
Tyrosine metabolism (map00350)
Phenylalanine metabolism (map00360)
Tryptophan metabolism (map00380)
Phenylalanine, tyrosine and tryptophan biosynthesis (map00400)
Urea cycle and metabolism of amino groups (map00220)

Metabolism of Other Amino Acids

beta-Alanine metabolism (map00410)
Taurine and hypotaurine metabolism (map00430)
Aminophosphonate metabolism (map00440)
Selenoamino acid metabolism (map00450)
Cyanoamino acid metabolism (map00460)
Glutamine and D-glutamate metabolism (map00471)
Arginine and D-ornithine metabolism (map00472)
D-Alanine metabolism (map00473)
Glutathione metabolism (map00480)

Glycan Biosynthesis and Metabolism

N-Glycan biosynthesis (map00510)
O-Glycan biosynthesis (map00512)
Ubiquinone biosynthesis (map00130)

Chondroitin / Heparan sulfate biosynthesis (map00532)
 Glycosaminoglycan degradation (map00531)
 Lipopolysaccharide biosynthesis (map00540)
 Peptidoglycan biosynthesis (map00550)
 Glycosylphosphatidylinositol(GPI)-anchor biosynthesis (map00563h)
 Glycosylphosphatidylinositol(GPI)-anchor biosynthesis (map00563y)
 Glycosphingolipid metabolism (map00600)
 Blood group glycolipid biosynthesis-lactoseries (map00601)
 Blood group glycolipid biosynthesis-neolactoseries (map00602)
 Globoside metabolism (map00603)
 Ganglioside biosynthesis (map00604)

Metabolism of Cofactors and Vitamins

Thiamine metabolism (map00730)
 Riboflavin metabolism (map00740)
 Vitamin B6 metabolism (map00750)
 Nicotinate and nicotinamide metabolism (map00760)
 Pantothenate and CoA biosynthesis (map00770)
 Biotin metabolism (map00780)
 Folate biosynthesis (map00790)
 One carbon pool by folate (map00670)
 Retinol metabolism (map00830)
 Porphyrin and chlorophyll metabolism (map00860)

9 Apendice 2

Pfams	No. Reaction Bio	% Reaction bio	No. Reaction kegg	%Reaction kegg	%Bio / %Kegg
PF00111	50	5,291005291	43	1,362916006	3,882121324
PF00355	43	4,55026455	22	0,697305864	6,525493025
PF00970	42	4,444444444	27	0,855784469	5,193415638
PF00175	42	4,444444444	37	1,17274168	3,78978979
PF00866	32	3,386243386	2	0,063391442	53,41798942
PF00848	28	2,962962963	2	0,063391442	46,74074074
PF00106	25	2,645502646	143	4,532488114	0,583675584
PF02332	21	2,222222222	1	0,031695721	70,11111111
PF00561	21	2,222222222	36	1,141045959	1,947530864
PF02406	19	2,010582011	1	0,031695721	63,43386243
PF00903	18	1,904761905	10	0,316957211	6,00952381
PF00171	17	1,798941799	67	2,123613312	0,847113638
PF04945	16	1,693121693	4	0,126782884	13,35449735
PF01494	13	1,375661376	44	1,394611727	0,986411736
PF07992	13	1,375661376	72	2,282091918	0,602807172
PF02746	12	1,26984127	11	0,348652932	3,642135642
PF00775	12	1,26984127	0	0	100
PF04663	12	1,26984127	0	0	100
PF01188	12	1,26984127	5	0,158478605	8,012698413
PF06099	12	1,26984127	0	0	100
PF04444	10	1,058201058	0	0	100
PF00037	9	0,952380952	39	1,236133122	0,77045177
PF01738	8	0,846560847	0	0	100
PF08240	7	0,740740741	57	1,806656101	0,410006498
PF01568	7	0,740740741	18	0,570522979	1,298353909
PF00378	7	0,740740741	35	1,109350238	0,667724868
PF00384	7	0,740740741	13	0,412044374	1,797720798
PF00107	7	0,740740741	81	2,567353407	0,288523091
PF01613	7	0,740740741	32	1,014263074	0,730324074
PF00487	6	0,634920635	9	0,28526149	2,225749559
PF00795	6	0,634920635	23	0,729001585	0,87094548
PF03241	5	0,529100529	7	0,221870048	2,38473167

PF02979	5	0,529100529	2	0,063391442	8,346560847
PF02770	5	0,529100529	28	0,88748019	0,596182918
PF00881	5	0,529100529	17	0,538827258	0,981948335
PF00296	5	0,529100529	20	0,633914422	0,834656085
PF01425	5	0,529100529	8	0,253565769	2,086640212
PF00501	5	0,529100529	59	1,870047544	0,282934266
PF02771	5	0,529100529	28	0,88748019	0,596182918
PF00465	5	0,529100529	27	0,855784469	0,618263766
PF02964	5	0,529100529	1	0,031695721	16,69312169
PF02211	4	0,423280423	2	0,063391442	6,677248677
PF00941	4	0,423280423	21	0,665610143	0,635928445
PF01315	4	0,423280423	10	0,316957211	1,335449735
PF01557	4	0,423280423	16	0,507131537	0,834656085
PF02900	4	0,423280423	4	0,126782884	3,338624339
PF02738	4	0,423280423	10	0,316957211	1,335449735
PF00701	4	0,423280423	9	0,28526149	1,483833039
PF01323	4	0,423280423	1	0,031695721	13,35449735
PF08028	4	0,423280423	17	0,538827258	0,785558668
PF01799	4	0,423280423	10	0,316957211	1,335449735
PF06234	4	0,423280423	0	0	100
PF00702	3	0,317460317	61	1,933438986	0,16419464
PF03450	3	0,317460317	10	0,316957211	1,001587302
PF01565	3	0,317460317	21	0,665610143	0,476946334
PF00043	3	0,317460317	12	0,380348653	0,834656085
PF01144	3	0,317460317	6	0,190174326	1,669312169
PF01979	3	0,317460317	17	0,538827258	0,589169001
PF01689	3	0,317460317	9	0,28526149	1,11287478
PF02461	3	0,317460317	1	0,031695721	10,01587302
PF00682	2	0,211640212	12	0,380348653	0,55643739
PF00732	2	0,211640212	20	0,633914422	0,333862434
PF01968	2	0,211640212	7	0,221870048	0,953892668
PF03243	2	0,211640212	0	0	100
PF02803	2	0,211640212	19	0,6022187	0,351434141
PF01011	2	0,211640212	9	0,28526149	0,74191652
PF04896	2	0,211640212	1	0,031695721	6,677248677
PF01575	2	0,211640212	101	3,201267829	0,066111373
PF02538	2	0,211640212	2	0,063391442	3,338624339

PF01361	2	0,211640212	0	0	100
PF00743	2	0,211640212	0	0	100
PF05199	2	0,211640212	7	0,221870048	0,953892668
PF02798	2	0,211640212	8	0,253565769	0,834656085
PF00724	2	0,211640212	14	0,443740095	0,476946334
PF05378	2	0,211640212	7	0,221870048	0,953892668
PF04744	2	0,211640212	1	0,031695721	6,677248677
PF02668	2	0,211640212	6	0,190174326	1,11287478
PF02913	2	0,211640212	15	0,475435816	0,445149912
PF00108	2	0,211640212	19	0,6022187	0,351434141
PF06192	2	0,211640212	0	0	100
PF07836	2	0,211640212	3	0,095087163	2,225749559
PF02769	1	0,105820106	7	0,221870048	0,476946334
PF00756	1	0,105820106	9	0,28526149	0,37095826
PF01883	1	0,105820106	1	0,031695721	3,338624339
PF04055	1	0,105820106	35	1,109350238	0,095389267
PF00291	1	0,105820106	26	0,824088748	0,128408628
PF06050	1	0,105820106	6	0,190174326	0,55643739
PF00462	1	0,105820106	1	0,031695721	3,338624339
PF00394	1	0,105820106	5	0,158478605	0,667724868
PF00725	1	0,105820106	26	0,824088748	0,128408628
PF07859	1	0,105820106	8	0,253565769	0,417328042
PF01228	1	0,105820106	6	0,190174326	0,55643739
PF00753	1	0,105820106	14	0,443740095	0,238473167
PF02852	1	0,105820106	10	0,316957211	0,333862434
PF04277	1	0,105820106	0	0	100
PF02776	1	0,105820106	31	0,982567353	0,107697559
PF07883	1	0,105820106	4	0,126782884	0,834656085
PF02962	1	0,105820106	1	0,031695721	3,338624339
PF03977	1	0,105820106	5	0,158478605	0,667724868
PF02560	1	0,105820106	1	0,031695721	3,338624339
PF08201	1	0,105820106	0	0	100
PF07858	1	0,105820106	0	0	100
PF02737	1	0,105820106	28	0,88748019	0,119236584
PF02241	1	0,105820106	0	0	100
PF00117	1	0,105820106	16	0,507131537	0,208664021
PF00441	1	0,105820106	29	0,919175911	0,115124977

PF01451	1	0,105820106	1	0,031695721	3,338624339
PF01855	1	0,105820106	3	0,095087163	1,11287478
PF03060	1	0,105820106	15	0,475435816	0,222574956
PF05440	1	0,105820106	0	0	100
PF01730	1	0,105820106	0	0	100
PF00120	1	0,105820106	4	0,126782884	0,834656085
PF02505	1	0,105820106	1	0,031695721	3,338624339
PF01493	1	0,105820106	3	0,095087163	1,11287478
PF01966	1	0,105820106	22	0,697305864	0,151755652
PF02663	1	0,105820106	1	0,031695721	3,338624339
PF01266	1	0,105820106	80	2,535657686	0,041732804
PF00699	1	0,105820106	1	0,031695721	3,338624339
PF03063	1	0,105820106	2	0,063391442	1,669312169
PF01913	1	0,105820106	1	0,031695721	3,338624339
PF00403	1	0,105820106	4	0,126782884	0,834656085
PF07746	1	0,105820106	0	0	100
PF07731	1	0,105820106	5	0,158478605	0,667724868
PF02894	1	0,105820106	23	0,729001585	0,14515758
PF03960	1	0,105820106	0	0	100
PF01774	1	0,105820106	0	0	100
PF07732	1	0,105820106	5	0,158478605	0,667724868
PF00449	1	0,105820106	1	0,031695721	3,338624339
PF02775	1	0,105820106	29	0,919175911	0,115124977
PF01118	1	0,105820106	7	0,221870048	0,476946334
PF01869	1	0,105820106	3	0,095087163	1,11287478
PF06243	1	0,105820106	0	0	100
PF04609	1	0,105820106	1	0,031695721	3,338624339
PF00596	1	0,105820106	27	0,855784469	0,123652753
PF08031	1	0,105820106	0	0	100
PF05721	1	0,105820106	0	0	100
PF02783	1	0,105820106	0	0	100
PF02552	1	0,105820106	0	0	100
PF01039	1	0,105820106	10	0,316957211	0,333862434
PF04206	1	0,105820106	1	0,031695721	3,338624339
PF01965	1	0,105820106	5	0,158478605	0,667724868
PF01977	1	0,105820106	11	0,348652932	0,303511304
PF05194	1	0,105820106	0	0	100

PF03201	1	0,105820106	2	0,063391442	1,669312169
PF04207	1	0,105820106	0	0	100
PF02745	1	0,105820106	1	0,031695721	3,338624339
PF01993	1	0,105820106	1	0,031695721	3,338624339
PF02826	1	0,105820106	37	1,17274168	0,09023309
PF02607	1	0,105820106	1	0,031695721	3,338624339
PF02289	1	0,105820106	1	0,031695721	3,338624339
PF01070	1	0,105820106	21	0,665610143	0,158982111
PF00547	1	0,105820106	1	0,031695721	3,338624339
PF05138	1	0,105820106	0	0	100
PF04211	1	0,105820106	0	0	100
PF02310	1	0,105820106	18	0,570522979	0,18547913
PF02901	1	0,105820106	2	0,063391442	1,669312169
PF02741	1	0,105820106	1	0,031695721	3,338624339
PF04209	1	0,105820106	1	0,031695721	3,338624339
PF02249	1	0,105820106	1	0,031695721	3,338624339
PF02814	1	0,105820106	0	0	100
PF02492	1	0,105820106	0	0	100
PF03061	1	0,105820106	11	0,348652932	0,303511304
PF02441	1	0,105820106	14	0,443740095	0,238473167
PF01593	1	0,105820106	56	1,77496038	0,059618292
PF02515	1	0,105820106	0	0	100
PF01569	1	0,105820106	17	0,538827258	0,196389667
PF00578	1	0,105820106	3	0,095087163	1,11287478
PF03328	1	0,105820106	10	0,316957211	0,333862434
PF00364	1	0,105820106	32	1,014263074	0,104332011
PF01042	1	0,105820106	0	0	100
PF00205	1	0,105820106	24	0,760697306	0,139109347
PF01408	1	0,105820106	27	0,855784469	0,123652753
PF01656	1	0,105820106	9	0,28526149	0,37095826

10 Apéndice 3

Cellular Processes

- Cell Motility
 - Bacterial chemotaxis
 - Flagellar assembly
 - Regulation of actin cytoskeleton
 - Bacterial motility proteins
- Cell Growth and Death
 - Cell cycle
 - Cell cycle - yeast
 - Apoptosis
 - p53 signaling pathway
- Cell Communication
 - Focal adhesion
 - Adherens junction
 - Tight junction
 - Gap junction
- Endocrine System
 - Insulin signaling pathway
 - Adipocytokine signaling pathway
 - PPAR signaling pathway
 - GnRH signaling pathway
 - Progesterone-mediated oocyte maturation
 - Melanogenesis
 - Renin-angiotensin system
- Immune System
 - Hematopoietic cell lineage
 - Complement and coagulation cascades
 - Toll-like receptor signaling pathway
 - Natural killer cell mediated cytotoxicity
 - Antigen processing and presentation
 - T cell receptor signaling pathway
 - B cell receptor signaling pathway
 - Fc epsilon RI signaling pathway
 - Leukocyte transendothelial migration
- Nervous System
 - Long-term potentiation
 - Long-term depression
- Sensory System

- Olfactory transduction
- Taste transduction
- Development
 - Dorso-ventral axis formation
 - Axon guidance
- Behavior
 - Circadian rhythm

Genetic Information Processing

- Transcription
 - RNA biosynthesis (illustration)
 - RNA polymerase
 - Basal transcription factors
 - Transcription factors
- Translation
 - Protein biosynthesis (illustration)
 - Ribosome
 - Aminoacyl-tRNA biosynthesis
 - Translation factors
 - Ribosome
- Folding, Sorting and Degradation
 - Protein export
 - Type II secretion system
 - Type III secretion system
 - Type IV secretion system
 - SNARE interactions in vesicular transport
 - Regulation of autophagy
 - Ubiquitin mediated proteolysis Major update!
 - Proteasome
 - SNAREs
- Replication and Repair
 - DNA polymerase

Human Diseases

- Cancers
 - Colorectal cancer
 - Pancreatic cancer
 - Glioma
 - Thyroid cancer
 - Acute myeloid leukemia
 - Chronic myeloid leukemia
 - Basal cell carcinoma

- Melanoma
- Renal cell carcinoma
- Bladder cancer
- Prostate cancer
- Endometrial cancer
- Small cell lung cancer
- Non-small cell lung cancer
- Neurodegenerative Diseases
 - Alzheimer's disease
 - Parkinson's disease
 - Amyotrophic lateral sclerosis (ALS)
 - Huntington's disease
 - Dentatorubropallidoluysian atrophy (DRPLA)
 - Prion diseases
- Metabolic Disorders
 - Type I diabetes mellitus
 - Type II diabetes mellitus
 - Maturity onset diabetes of the young
- Infectious Diseases
 - Cholera
 - Epithelial cell signaling in Helicobacter pylori infection
 - Pathogenic Escherichia coli infection

Environmental Information Processing

- Membrane Transport
 - ABC transporters
 - Phosphotransferase system (PTS)
 - Transporters (+diseases)
- Signal Transduction
 - Two-component system
 - MAPK signaling pathway
 - ErbB signaling pathway
 - Wnt signaling pathway
 - Notch signaling pathway
 - Hedgehog signaling pathway
 - TGF-beta signaling pathway
 - VEGF signaling pathway
 - Jak-STAT signaling pathway
 - Calcium signaling pathway
 - Phosphatidylinositol signaling system
 - mTOR signaling pathway Two-component system

- Signaling Molecules and Interaction
 - Neuroactive ligand-receptor interaction
 - Cytokine-cytokine receptor interaction
 - ECM-receptor interaction
 - Cell adhesion molecules (CAMs)
 - Receptors and channels (+diseases)
 - Cytokines
 - Cell adhesion molecules (CAMs)
 - CAM ligands
 - CD molecules
 - GTP-binding proteins

11 Apéndice 4

11.1 BioNeMo: Molecular Information on biodegradation Metabolism.

Almudena Trigo, Guillermo Carbajosa, Alfonso Valencia and Ildefonso Cases

Spanish Center for Cancer Research, Madrid, Spain

Abstract.

Microorganisms play a critical role in the elimination of the numerous toxic compounds that accumulate in the environment as consequence of human activities. Decades of biochemical studies have produced a wealth of knowledge about this unique metabolism, which only recently has started to be categorized and stored in structured databases, such as the Metarouter, or the University of Minnesota Biocatalysis/Biodegradation Database (UMBBD). The formalization of this biochemical knowledge has allowed the first attempts to approach this process from a Systems Biology perspective. However, the absence of information on the proteins present in this network has precluded further systematic studies. To fill this gap, we have incorporated protein information for many of the reactions in the biodegradation network by literature mining, customized database searches and manual curation, including sequence, protein domains, structures, genetic organization and transcriptional regulation. This database, called BioNeMo, is available through a web server, but it is also possible to download it and install it locally. To facilitate the access to the underlying relational database, an Object-Oriented Perl API is also provided. Examples of use of both accession methods are provided.

11.2 Characterization of Enzyme Properties at the Global Biodegradation Network

Almudena Trigo, Alfonso Valencia and Ildefonso Cases

Spanish Center for Cancer Research, Madrid, Spain

Abstract.

Bioremediation, the exploitation of microorganisms for removing pollutants from the environment, is an area of great importance due to the high amount of harmful compounds released in the environment, which has been still scarcely addressed from a Systems perspective. Previous works studied the Global Biodegradation Network over a classical metabolic network formulation, where the compounds are treated as nodes connected by reactions. In this work, we have constructed an alternative, reaction-centric view of the network, in which reactions are nodes that are connected if the product of one can be the substrate of the other. In this way, we can now focus in the biological entities (proteins that perform the reactions) instead of chemical entities (the compounds). For each reaction, we have associated detailed and manually curated sequence, functional and structural information. The mapping of this information provides, not only an overview of the general knowledge of the degradation network, but also new insights into its structure, function and evolution. The analysis described in this work reveals a linear structure of the network, with an increased complexity and promiscuity and decreased sequence variation of the enzymes in the periphery of the network. Possible interpretations of these observations in the context of the dynamics, evolution and evolvability of the Biodegradation network are discussed.

11.3 Origin, Evolution and Assembly of Enzymes and Pathways of the Global Biodegradation Network

Almudena Trigo, Alfonso Valencia and Ildefonso Cases

Spanish Center for Cancer Research, Madrid, Spain

Abstract.

The microbial degradation of toxic compounds, product of extensive human activities, the biodegradation metabolism, is key for the regeneration of polluted sites. Many of these compounds have started to accumulate in the biosphere only recently, so it has been hypothesized that the microbial activities for their mineralization have also originated recently, and is still under rapid evolution. So far, the evolution of only a few individual biodegradation enzymes or pathways has been analyzed, and these studies have been focused mainly in the origin of the catalytic activities. In this work, we present the first systematic analysis of the homology relation between the enzymes of the biodegradation network and those of the central metabolism, and the relations them and other enzymes of the global biodegradation network. Our results suggest that most of the biodegradation complement evolved from a small set of ancient enzymes, and over history, has expanded by duplication and shuffling into new pathways. Interestingly, different mechanism for functional expansion seems to have operated in different areas of the network, suggesting that different selective forces operate on these areas.